Shakil Mahmud • Research Monograph

SHAKIL MAHMUD EDITORIAL & RESEARCH MONOGRAPHS

মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব

পিএসি শিখন ও ভিসি ডাইমেনশন থেকে লস জ্যামিতি, উত্তল অপ্টিমাইজেশন, এসভিএম মার্জিন, কার্নেল পদ্ধতি, বুস্টিং এবং বায়েসীয় গ্রাফিক্যাল ইনফারেন্সের ২০ খণ্ডের এক প্রামাণ্য তাত্ত্বিক মহাকাব্য।

SHAKIL MAHMUD

গবেষক ও লেখক

SHAKIL MAHMUD EDITORIAL & RESEARCH SERIES

প্রথম সংকলিত সংস্করণ • ঢাকা ও সিঙ্গাপুর • ২০২৬

মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব

পিএসি শিখন ও ভিসি ডাইমেনশন থেকে লস জ্যামিতি, উত্তল অপ্টিমাইজেশন, এসভিএম মার্জিন, কার্নেল পদ্ধতি, বুস্টিং এবং বায়েসীয় গ্রাফিক্যাল ইনফারেন্সের ২০ খণ্ডের এক প্রামাণ্য তাত্ত্বিক মহাকাব্য।

LIBRARY & CATALOGING-IN-PUBLICATION DATA

Title: মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব
Author: Shakil Mahmud
Series: Shakil Mahmud Editorial & Research Series (Monograph Collection)
Subjects: Science -- History -- Philosophy -- Mathematics -- Civilization
Published: Dhaka & Singapore, 2026
Format: Digital Monograph Offprint (PDF) • Complete 20-Part Collected Edition
Permanent URI: https://shakilmahmud.com/series/machine-learning/

প্রস্তাবিত সাইটেশন:
Mahmud, Shakil. (2026). *মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব: The Complete 20-Part Collected Monograph*. Shakil Mahmud Research Series. https://shakilmahmud.com/series/machine-learning/

Open Scholarly Access: This collected monograph is published under Open Scholarly Access for personal study, academic citation, university pedagogy, and non-commercial research.

© 2026 Shakil Mahmud. All rights reserved. Typeset and published at Dhaka and Singapore.

Shakil Mahmud • Research Monograph

তাত্ত্বিক ভূমিকা

ভূমিকা: জ্ঞানের একীভূত রূপরেখা

❖

পরিসংখ্যানিক শিখন তত্ত্ব হলো সেই গাণিতিক ভিত্তি যা ব্যাখ্যা করে কীভাবে সসীম উপাত্ত থেকে সার্বজনীন ভবিষ্যদ্বাণীমূলক মডেল তৈরি সম্ভব। ভ্যাপনিক-চারভোনেনকিস (ভিসি) ডাইমেনশন ও রাডেম্যাকার জটিলতা থেকে রিপ্রোডিউসিং কার্নেল হিলবার্ট স্পেস (আরকেএইচএস), রেগুলারাইজড রিস্ক মিনিমাইজেশন, ফাংশনাল স্পেসে বুস্টিং এবং সসীম নমুনায় কনফরমাল প্রেডিকশন—বিশ খণ্ডে সংকলিত এই গবেষণা মনোগ্রাফ আধুনিক মেশিন লার্নিংয়ের অন্তর্নিহিত জ্যামিতি, সম্ভাব্যতা ও অপ্টিমাইজেশন তত্ত্বের নিখুঁত গাণিতিক বিশ্লেষণ উপস্থাপন করে।

এই গবেষণামূলক কাজের উদ্দেশ্য কেবল তথ্য উপস্থাপন নয়, বরং চিন্তার গভীর বিকাশকে পাঠকের সামনে উন্মোচন করা। প্রতিটি পরিচ্ছদ এক একটি যুগান্তকারী বৌদ্ধিক সংকটের সমাধান হিসেবে কীভাবে আবির্ভূত হয়েছে, তা এখানে পদ্ধতিগতভাবে লিপিবদ্ধ করা হলো।

— Shakil Mahmud
Dhaka & Singapore, 2026

অধ্যায় সূচি

সূচিপত্র

প্রথম পর্যায়: সম্ভাব্যতা, অপ্টিমাইজেশন ও লিনিয়ার মডেল

খণ্ড ১ আগমনের জ্যামিতি — এম্পিরিক্যাল রিস্ক মিনিমাইজেশন, ইনডাক্টিভ বায়াস এবং জেনারেলাইজেশন — 606 শব্দ
আগমনের জ্যামিতি — এম্পিরিক্যাল রিস্ক মিনিমাইজেশন, ইনডাক্টিভ বায়াস এবং জেনারেলাইজেশন

খণ্ড ২ পরিসংখ্যানিক শিখন তত্ত্ব — ভ্যালিয়ান্টের পিএসি ফ্রেমওয়ার্ক, নমুনা জটিলতা এবং ওকামের ক্ষুর — 524 শব্দ
পরিসংখ্যানিক শিখন তত্ত্ব — ভ্যালিয়ান্টের পিএসি ফ্রেমওয়ার্ক, নমুনা জটিলতা এবং ওকামের ক্ষুর

খণ্ড ৩ মডেল সক্ষমতা ও ভিসি ডাইমেনশন — রাডমাখার জটিলতা, গ্রোথ ফাংশন এবং শ্যাটারিং গুণাঙ্ক — 504 শব্দ
মডেল সক্ষমতা ও ভিসি ডাইমেনশন — রাডমাখার জটিলতা, গ্রোথ ফাংশন এবং শ্যাটারিং গুণাঙ্ক

খণ্ড ৪ বায়াস-ভ্যারিয়েন্স ভারসাম্য — প্রত্যাশিত পূর্বাভাস ত্রুটির বিশ্লেষণ ও ওভারফিটিং গতিবিজ্ঞান — 418 শব্দ
বায়াস-ভ্যারিয়েন্স ভারসাম্য — প্রত্যাশিত পূর্বাভাস ত্রুটির বিশ্লেষণ ও ওভারফিটিং গতিবিজ্ঞান

দ্বিতীয় পর্যায়: পরিসংখ্যানিক শিখন তত্ত্ব ও নন-লিনিয়ার মার্জিন

খণ্ড ৫ বহু-মাত্রিক উত্তল অপ্টিমাইজেশন — গ্রেডিয়েন্ট ডিসেন্ট, সাব-গ্রেডিয়েন্ট এবং রূপান্তর গতিবিজ্ঞান — 502 শব্দ
বহু-মাত্রিক উত্তল অপ্টিমাইজেশন — গ্রেডিয়েন্ট ডিসেন্ট, সাব-গ্রেডিয়েন্ট এবং রূপান্তর গতিবিজ্ঞান

খণ্ড ৬ স্টোকাস্টিক অপ্টিমাইজেশন — মিনি-ব্যাচ এসজিডি, মোমেন্টাম এবং অ্যাডাপ্টিভ মোমেন্টস (অ্যাডাম) — 420 শব্দ
স্টোকাস্টিক অপ্টিমাইজেশন — মিনি-ব্যাচ এসজিডি, মোমেন্টাম এবং অ্যাডাপ্টিভ মোমেন্টস (অ্যাডাম)

খণ্ড ৭ রৈখিক বিভাজক ও সার্বজনীন রৈখিক মডেল — ফিশারের এলডিএ এবং এক্সপোনেনশিয়াল ফ্যামিলি — 412 শব্দ
রৈখিক বিভাজক ও সার্বজনীন রৈখিক মডেল — ফিশারের এলডিএ এবং এক্সপোনেনশিয়াল ফ্যামিলি

খণ্ড ৮ মার্জিনের জ্যামিতি — সাপোর্ট ভেক্টর মেশিন, কেকেটি শর্তাবলি এবং সাপোর্ট ভেক্টর — 529 শব্দ
মার্জিনের জ্যামিতি — সাপোর্ট ভেক্টর মেশিন, কেকেটি শর্তাবলি এবং সাপোর্ট ভেক্টর

তৃতীয় পর্যায়: লাগ্রাঞ্জ দ্বৈততা, এনসেম্বল ও বুস্টিং

খণ্ড ৯ রিপ্রোডিউসিং কার্নেল হিলবার্ট স্পেস — মার্সারের উপপাদ্য এবং অসীম-মাত্রিক কার্নেল ট্রিক — 495 শব্দ
রিপ্রোডিউসিং কার্নেল হিলবার্ট স্পেস — মার্সারের উপপাদ্য এবং অসীম-মাত্রিক কার্নেল ট্রিক

খণ্ড ১০ স্পার্সিটি ও রেগুলারাইজেশন মেকানিক্স — এল১ ল্যাসো, এল২ রিজ এবং সাব-ডিফারেনশিয়াল জ্যামিতি — 483 শব্দ
স্পার্সিটি ও রেগুলারাইজেশন মেকানিক্স — এল১ ল্যাসো, এল২ রিজ এবং সাব-ডিফারেনশিয়াল জ্যামিতি

খণ্ড ১১ ডিসিশন ট্রি ইন্ডাকশন ও রিকার্সিভ পার্টিশনিং — ইনফরমেশন গেইন, জিনি ইমপিউরিটি এবং প্রুনিং — 432 শব্দ
ডিসিশন ট্রি ইন্ডাকশন ও রিকার্সিভ পার্টিশনিং — ইনফরমেশন গেইন, জিনি ইমপিউরিটি এবং প্রুনিং

খণ্ড ১২ এনসেম্বল লার্নিং ও ভ্যারিয়েন্স রিডাকশন — বুটস্ট্র্যাপ অ্যাগ্রিগেশন, ওওবি বাউন্ড এবং র‍্যান্ডম ফরেস্ট — 511 শব্দ
এনসেম্বল লার্নিং ও ভ্যারিয়েন্স রিডাকশন — বুটস্ট্র্যাপ অ্যাগ্রিগেশন, ওওবি বাউন্ড এবং র‍্যান্ডম ফরেস্ট

খণ্ড ১৩ ফাংশনাল স্পেসে বুস্টিং — অ্যাডাবুস্ট, মার্জিন বাউন্ড এবং এক্সজিবিউস্ট — 468 শব্দ
ফাংশনাল স্পেসে বুস্টিং — অ্যাডাবুস্ট, মার্জিন বাউন্ড এবং এক্সজিবিউস্ট

খণ্ড ১৪ সম্ভাবনাময় গ্রাফিক্যাল মডেল — বায়েসিয়ান নেটওয়ার্ক, এমআরএফ এবং ডি-সেপারেশন — 422 শব্দ
সম্ভাবনাময় গ্রাফিক্যাল মডেল — বায়েসিয়ান নেটওয়ার্ক, এমআরএফ এবং ডি-সেপারেশন

Shakil Mahmud • Research Monograph

চতুর্থ পর্যায়: নন-প্যারামেট্রিক, ম্যানিফোল্ড ও অনিশ্চয়তা গ্যারান্টি

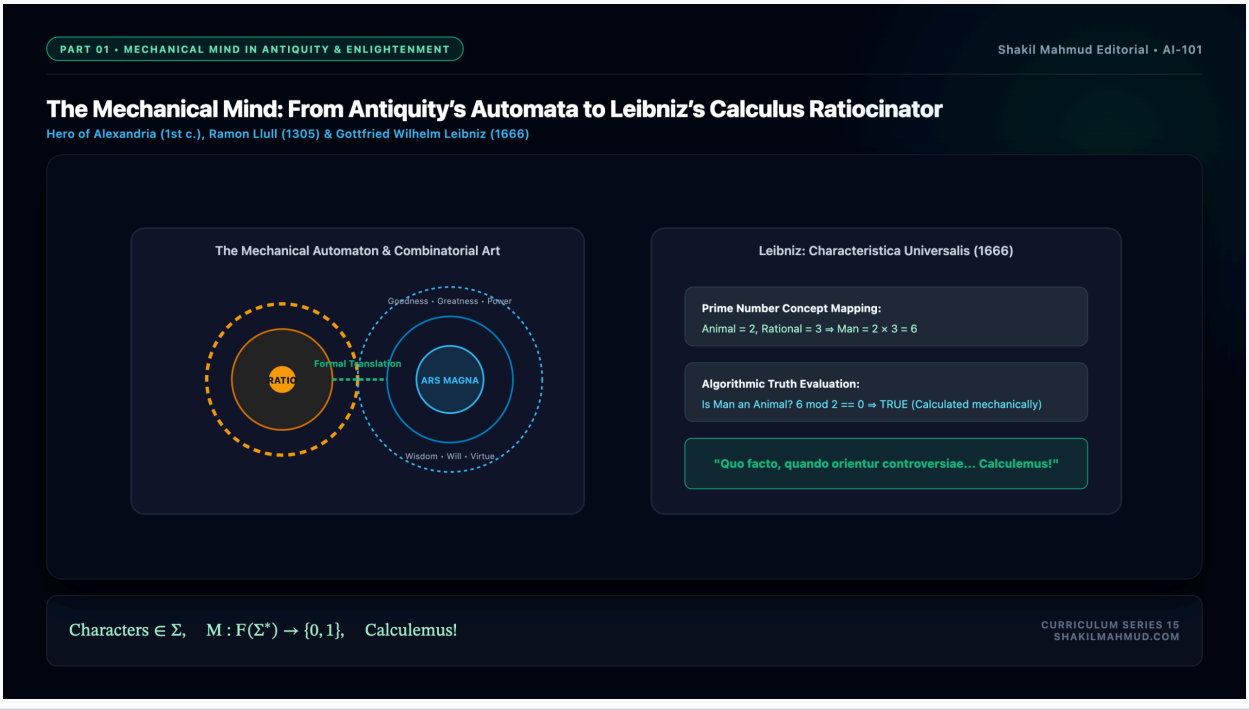
খণ্ড ১৫	**লুকায়িত ভেরিয়েবল মডেল — জেনসেনের উপপাদ্য এবং ইএম অ্যালগরিদম** *লুকায়িত ভেরিয়েবল মডেল — জেনসেনের উপপাদ্য এবং ইএম অ্যালগরিদম*	413 শব্দ
খণ্ড ১৬	**আসন্ন অনুমান ও ভ্যারিয়েশনাল বায়েস — মিন-ফিল্ড রূপরেখা, কেএল ডাইভারজেন্স এবং ইএলবিও** *আসন্ন অনুমান ও ভ্যারিয়েশনাল বায়েস — মিন-ফিল্ড রূপরেখা, কেএল ডাইভারজেন্স এবং ইএলবিও*	395 শব্দ
খণ্ড ১৭	**গাউসিয়ান প্রসেস ও বায়েসিয়ান নন-প্যারামেট্রিক তত্ত্ব — অসীম মাত্রার ফাংশন স্পেস** *গাউসিয়ান প্রসেস ও বায়েসিয়ান নন-প্যারামেট্রিক তত্ত্ব — অসীম মাত্রার ফাংশন স্পেস*	381 শব্দ
খণ্ড ১৮	**ডাইমেনশনালিটি রিডাকশন ও ম্যানিফোল্ড লার্নিং — পিসিএ থেকে ইউম্যাপ** *ডাইমেনশনালিটি রিডাকশন ও ম্যানিফোল্ড লার্নিং — পিসিএ থেকে ইউম্যাপ*	389 শব্দ
খণ্ড ১৯	**সেমি-সুপারভাইজড ও অ্যাক্টিভ লার্নিং — স্বল্প-লেবেল রেজিমের ইনফরমেশন জিওমেট্রি** *সেমি-সুপারভাইজড ও অ্যাক্টিভ লার্নিং — স্বল্প-লেবেল রেজিমের ইনফরমেশন জিওমেট্রি*	369 শব্দ
খণ্ড ২০	**মডেল ক্যালিব্রেশন ও কনফরমাল প্রেডিকশন — সসীম নমুনায় নিশ্চিত পূর্বাভাসের গাণিতিক গ্যারান্টি** *মডেল ক্যালিব্রেশন ও কনফরমাল প্রেডিকশন — সসীম নমুনায় নিশ্চিত পূর্বাভাসের গাণিতিক গ্যারান্টি*	464 শব্দ

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • প্রথম খণ্ড

আগমনের জ্যামিতি — এম্পিরিক্যাল রিস্ক মিনিমাইজেশন, ইনডাক্টিভ বায়াস এবং জেনারেলাইজেশন

আগমনের জ্যামিতি — এম্পিরিক্যাল রিস্ক মিনিমাইজেশন, ইনডাক্টিভ বায়াস এবং জেনারেলাইজেশন

606 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১: *যান্ত্রিক অটোমাটা ও রামোন লুলের কম্বিনেটোরিয়াল আর্স ম্যাগনা এবং গটফ্রিড ভিলহেম লাইবনিজের ক্যারেক্টারিস্টিকা ইউনিভার্সালিস ও ক্যালকুলাস র‍্যাটিওসিনেটরের রূপরেখা।*

মেশিন লার্নিংয়ের দার্শনিক কেন্দ্রে অবস্থান করে এক প্রাচীন জিজ্ঞাসা—আগমনের সমস্যা (Problem of Induction): অতীতের সসীম পর্যবেক্ষণের ওপর ভিত্তি করে কীভাবে অদেখা ভবিষ্যতের ঘটনা সম্পর্কে গাণিতিকভাবে সুদৃঢ় সিদ্ধান্ত নেওয়া সম্ভব? দার্শনিক ডেভিড হিউম যেখানে সংশয় ব্যক্ত করেছিলেন, আধুনিক **পরিসংখ্যানিক শিখন তত্ত্ব (Statistical Learning Theory)** সেই দার্শনিক ধাঁধাকে ফাংশন অপ্টিমাইজেশন, সম্ভাব্যতা পরিমাপ এবং জ্যামিতিক বাউন্ডের সুনির্দিষ্ট বিজ্ঞানে রূপান্তরিত করেছে। শিখন হলো সীমাবদ্ধ সম্ভাবনার পরিসরে এক সুশৃঙ্খল গাণিতিক আগমন।

১. গাণিতিক কাঠামো: ডেটা ডিস্ট্রিবিউশন ও ট্রু রিস্ক

ধরা যাক $mathcalX subseteq mathbbR^d$ হলো ইনপুট বা ইনস্ট্যান্স স্পেস এবং $mathcalY$ হলো টার্গেট লেবেল স্পেস (যেমন বাইনারি ক্লাসিফিকেশনের ক্ষেত্রে $mathcalY = -1, +1$ বা রিগ্রেশনের জন্য $mathcalY subseteq mathbbR$)। প্রকৃতি একটি অজ্ঞাত কিন্তু নির্দিষ্ট সম্ভাব্যতা বণ্টন $mathcalD$-এর মাধ্যমে ডেটা জোড়া $(x, y) in mathcalX times mathcalY$ তৈরি করে:

$$(x, y) sim mathcalD(x, y) = mathcalD(x) mathcalD(y mid x)$$

একজন শিক্ষার্থীর কাছে এই সামগ্রিক ডিস্ট্রিবিউশন $mathcalD$-তে সরাসরি প্রবেশাধিকার থাকে না। এর পরিবর্তে সে $mathcalD$ থেকে স্বাধীন ও অভিন্নভাবে বণ্টিত (i.i.d.) n আকারের একটি ট্রেনিং সেট S পর্যবেক্ষণ করে:

Shakil Mahmud • Research Monograph

$$S = (x_1, y_1), (x_2, y_2), dots, (x_n, y_n) simmathcalD^n$$

একটি লস ফাংশন $L : mathcalY times mathcalY to mathbbR_{ge0}$ পূর্বাভাষ $haty = h(x)$ এবং আসল মান y-এর পার্থক্য পরিমাপ করে। শিখনের চূড়ান্ত লক্ষ্য হলো এমন একটি হাইপোথিসিস $hin mathcalH$ খুঁজে বের করা যা **প্রকৃত প্রত্যাশিত ঝুঁকি বা ট্রু রিস্ক (True Expected Risk)** সর্বনিম্ন করে:

$$R(h) = mathbbE_{(x,y)simmathcalD}[L(h(x), y)] = int_{mathcalX times mathcalY} L(h(x), y), dmathcalD(x, y)$$

২. এম্পিরিক্যাল রিস্ক মিনিমাইজেশন (ERM) নীতি

যেহেতু ডেটার আসল বিন্যাস $mathcalD$ অজ্ঞাত, তাই সরাসরি ট্রু রিস্কের ইন্টিগ্রাল $R(h)$ গণনা করা অসম্ভব। ভ্লাদিমির ভ্যাপনিক এর একটি বাস্তবমুখী বিকল্প প্রস্তাব করেন যা **এম্পিরিক্যাল রিস্ক মিনিমাইজেশন (ERM)** নামে পরিচিত। এখানে সমগ্র ডিস্ট্রিবিউশনের বদলে ট্রেনিং ডেটার ওপর গড় লস গণনা করা হয়:

$$R_{textemp}(h; S) = frac1nsum_{i=1}^{n} L(h(x_i), y_i)$$

ইআরএম নিয়ম অনুসারে সেই হাইপোথিসিস $hath_S in mathcalH$-কে নির্বাচন করা হয় যা এই নমুনা গড় লসকে সর্বনিম্ন করে:

$$hath_S = argmin_{hinmathcalH} R_{textemp}(h; S)$$

বৃহৎ সংখ্যার সূত্র (Law of Large Numbers) অনুযায়ী, যেকোনো একটি নির্দিষ্ট h-এর জন্য নমুনা সংখ্যা $ntoinfty$ হলে এম্পিরিক্যাল রিস্ক ট্রু রিস্কের সমান হয়ে যায়। কিন্তু শিখনের আসল সংকট হলো—নির্বাচিত মডেল $hath_S$ কিন্তু পূর্বনির্ধারিত নয়; এটি স্বয়ং ডেটাসেট S-এর ওপর নির্ভর করে নির্বাচিত হয়েছে।

৩. অবাধ হাইপোথিসিস স্পেসের ব্যর্থতা: ওভারফিটিং ও মুখস্থকরণ

যদি হাইপোথিসিস স্পেস $mathcalH$-এর ওপর কোনো বিধিনিষেধ না থাকে এবং এটি যেকোনো ফাংশন গ্রহণ করতে পারে, তবে মডেলটি খুব সহজেই ডেটা মুখস্থ করে ফেলতে পারে:

h_{text{mem}}(x) = begin{cases} y_i & text{যদি } x = x_i in S \ text{যেকোনো মান} & text{অন্যথায়} end{cases}

এই মডেলের ক্ষেত্রে ট্রেনিং লস হবে শূন্য ($R_{textemp} = 0$)। কিন্তু ট্রেনিং সেটের বাইরে নতুন যেকোনো ডেটায় এর ভুল হবে চরম। এই ব্যবধানটিকেই বলা হয় **জেনারেলাইজেশন গ্যাপ (Generalization Gap)**। মডেলের ধারণক্ষমতা অতিরিক্ত বেশি হলে এম্পিরিক্যাল রিস্ক কোনোভাবেই আসল ঝুঁকির নির্দেশক হতে পারে না—যাকে আমরা বলি ওভারফিটিং।

৪. ইনডাক্টিভ বায়াস (Inductive Bias)-এর অপরিহার্যতা

আধুনিক মেশিন লার্নিংয়ের সবচেয়ে কেন্দ্রীয় আবিষ্কার হলো: **পূর্বানুমান বা ইনডাক্টিভ বায়াস ছাড়া কোনো শিখন সম্ভব নয়**। ওলপার্ট ও ম্যাকরেডির বিখ্যাত *নো ফ্রি লাঞ্চ থিওরেম (No Free Lunch Theorem)* প্রমাণ করে যে, সম্ভাব্য সকল ডেটা ডিস্ট্রিবিউশনের ওপর গড় করলে পৃথিবীর সকল শিখন অ্যালগরিদমের কার্যক্ষমতা সমান। অর্থাৎ, নতুন তথ্যে সফল হতে হলে অ্যালগরিদমকে ডেটা দেখার আগেই কিছু নির্দিষ্ট সমাধানকে অন্যদের চেয়ে বেশি প্রাধান্য দিতে হবে।

এই ইনডাক্টিভ বায়াস প্রতিষ্ঠিত হয় বিভিন্ন গাণিতিক কাঠামোর মাধ্যমে:

- **হাইপোথিসিস সীমাবদ্ধকরণ:** সমাধানকে কেবল সরলরেখা, কম মাত্রার বহুপদী বা নির্দিষ্ট গভীরতার ট্রিতে সীমাবদ্ধ রাখা।
- **মসৃণতা ও মার্জিনের শর্ত:** কাছাকাছি ডেটা পয়েন্টের জন্য একই লেবেল দেওয়ার নীতি বা ডিসিশন বাউন্ডারি থেকে দূরত্ব (Margin) সর্বোচ্চ করা।
- **নর্ম রেগুলারাইজেশন:** অপ্রয়োজনীয় জটিলতা কমাতে প্যারামিটারের ওপর L_1 বা L_2 পেনাল্টি আরোপ করা।

সংক্ষেপে, মেশিন লার্নিং হলো এই ইনডাক্টিভ বিধিনিষেধগুলোর এমন এক নিখুঁত জ্যামিতিক বিন্যাস যা সীমিত নমুনা তথ্যের ভেতরেও বিশ্বস্ত সার্বজনীন জ্ঞান উন্মোচন করতে সক্ষম।

সাইটেশন: মাহমুদ, শাকিল (2026)। "আগমনের জ্যামিতি — এম্পিরিক্যাল রিস্ক মিনিমাইজেশন, ইনডাক্টিভ বায়াস এবং জেনারেলাইজেশন"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-01-geometry-of-induction-erm-generalization-bn/

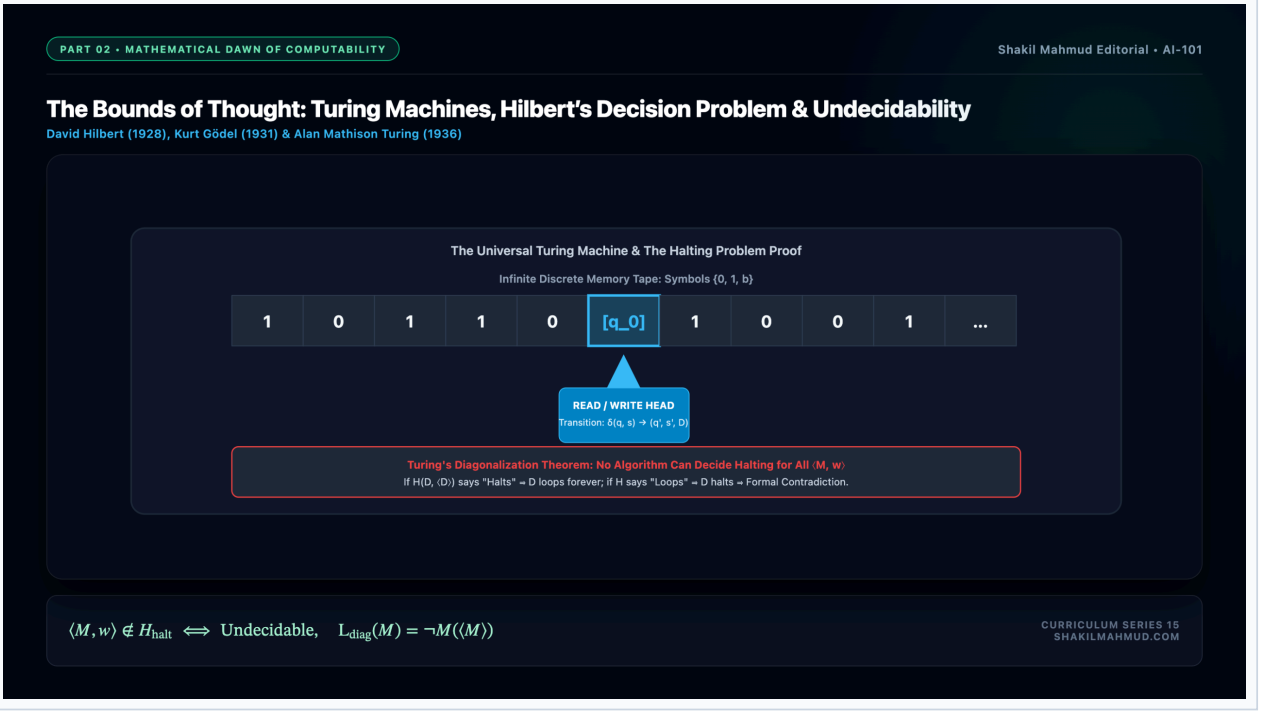
Shakil Mahmud • Research Monograph

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • দ্বিতীয় খণ্ড

পরিসংখ্যানিক শিখন তত্ত্ব — ভ্যালিয়ান্টের পিএসি ফ্রেমওয়ার্ক, নমুনা জটিলতা এবং ওকামের ক্ষুর

পরিসংখ্যানিক শিখন তত্ত্ব — ভ্যালিয়ান্টের পিএসি ফ্রেমওয়ার্ক, নমুনা জটিলতা এবং ওকামের ক্ষুর

524 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ২: *ইউনিভার্সাল টুরিং মেশিনের ডিসক্রিট টেপ আর্কিটেকচার এবং হল্টিং প্রবলেমের অনির্ণেয়তা প্রমাণকারী ডায়াগোনালাইজেশন মেকানিজ্ম।*

১৯৮৪ সালে কম্পিউটার বিজ্ঞানী লেসলি ভ্যালিয়ান্ট একটি যুগান্তকারী গবেষণাপত্র প্রকাশ করেন যা কৃত্রিম বুদ্ধিমত্তার তাত্ত্বিক ভিত্তি স্থাপন করেছিল: *"A Theory of the Learnable"*। ভ্যালিয়ান্ট শিখনের অস্পষ্ট অনুমানগুলোকে সরিয়ে এক সুনির্দিষ্ট, ডিস্ট্রিবিউশন-মুক্ত গাণিতিক কাঠামোর জন্ম দেন, যার নাম **পিএসি শিখন (PAC Learning — Probably Approximately Correct)**। প্রথমবারের মতো শিখন কেবল ট্রায়াল-অ্যান্ড-এরর বা কারিগরি কৌশল না থেকে পলিনমিয়াল নমুনা জটিলতা এবং নির্ভুলতার সুনির্দিষ্ট বিজ্ঞানে পরিণত হয়।

১. পিএসি-র দর্শন: কেন "Probably" এবং কেন "Approximately"?

যেকোনো বাস্তব শিখন প্রক্রিয়ায় দুটি অপরিবর্তনীয় সত্য বিদ্যমান:

1. **আনুমানিক নির্ভুলতা (*epsilon* — Approximately Correct):** যেহেতু ট্রেনিং ডেটা হলো অসীম বাস্তবতার এক ক্ষুদ্র নমুনা, তাই নতুন ডেটায় শতভাগ নির্ভুলতা ($R(h) = 0$) দাবি করা গাণিতিকভাবে অসম্ভব। অ্যালগরিদম কেবল এই নিশ্চয়তা দিতে পারে যে তার ভুল একটি ক্ষুদ্র সীমা epsilon > 0-এর মধ্যে থাকবে:

$$R(h) leepsilon$$

2. **উচ্চ সম্ভাব্যতা (1 – delta — Probably):** ট্রেনিং ডেটাসেট $Ssimmathcal D^n$ দৈবভাবে সংগৃহীত হয়। তাত্ত্বিকভাবে সবসময় একটি সম্ভাবনা থাকে যে সংগৃহীত নমুনাগুলো চরম বিভ্রান্তিকর ছিল। তাই আমরা নিশ্চিত করি যে শিখন ব্যর্থ হওয়ার সম্ভাবনা $deltain(0,1)$-এর কম থাকবে, অর্থাৎ সফলতার সম্ভাব্যতা হবে অন্তত 1 – delta:

mathbb{P}_{S sim mathcal{D}^n}(R(h_S) le epsilon) ge 1 – delta

২. পিএসি শিখনযোগ্যতার আনুষ্ঠানিক সংজ্ঞা

একটি কনসেপ্ট ক্লাস $mathcalC$-কে পিএসি-শিখনযোগ্য বলা হবে যদি এমন একটি অ্যালগরিদম $mathcalA$ থাকে যা যেকোনো অজানা বণ্টন $mathcalD$ এবং যেকোনো কাঙ্ক্ষিত নির্ভুলতা $epsilon$ ও আত্মবিশ্বাস 1 – delta-এর জন্য:

n আকারের ট্রেনিং নমুনা S গ্রহণ করে এমন একটি হাইপোথিসিস $hinmathcalH$ তৈরি করতে পারে যা অন্তত 1 – delta নিশ্চয়তায় $epsilon$-এর কম ভুল প্রদর্শন করে:

mathbb{P}_{S sim mathcal{D}^n}left[R(h) le epsilon right] ge 1 – delta

সবচেয়ে গুরুত্বপূর্ণ শর্ত হলো—এই অ্যালগরিদমের প্রয়োজনীয় নমুনা সংখ্যা $n = m(epsilon, delta)$ এবং রানটাইমকে $1/epsilon$, $1/delta$ এবং ডাইমেনশন d-এর পলিনমিয়াল বহুপদী হতে হবে।

৩. সসীম হাইপোথিসিস স্পেসে নমুনা জটিলতা (Sample Complexity)

ধরা যাক হাইপোথিসিস স্পেস $mathcalH$ সসীম এবং এর উপাদান সংখ্যা $|mathcalH|$। ডেটাসেটে এমন একটি হাইপোথিসিস আছে যার আসল ভুল অনেক বেশি (R(h) > epsilon), কিন্তু ট্রেনিং ডেটায় দৈবক্রমে সে কোনো ভুল করেনি ($R_{textemp}(h) = 0$)।

একটি খারাপ হাইপোথিসিসের একটি দৈব নমুনাকে সঠিকভাবে অনুমান করার সম্ভাবনা সর্বোচ্চ 1 – epsilon। সুতরাং পরপর n টি স্বাধীন নমুনায় তার ভুল না করার সম্ভাবনা হলো:

mathbb{P}(R_{text{emp}}(h) = 0 mid R(h) > epsilon) le (1 – epsilon)^n le e^{-nepsilon}

এখন স্পেসের সকল খারাপ হাইপোথিসিসের ওপর **ইউনিয়ন বাউন্ড (Union Bound)** প্রয়োগ করলে পাই:

mathbb{P}left(exists h in mathcal{H} : R_{text{emp}}(h) = 0 text{ এবং } R(h) > epsilonright) le |mathcal{H}| e^{-nepsilon}

আমরা চাই এই বিপর্যয়ের সম্ভাবনা যেন সর্বোচ্চ $delta$ হয়:

$$|mathcalH|e^{-nepsilon}ledeltaimpliesngefrac1epsilonleft(ln|mathcalH| + lnfrac1deltaright)$$

এটিই হলো বিখ্যাত **সসীম পিএসি স্যাম্পল কমপ্লেক্সিটি সূত্র**।

৪. ওকামের ক্ষুরের গাণিতিক প্রমাণ

চতুর্দশ শতাব্দীর দার্শনিক উইলিয়াম অব ওকামের সুবিখ্যাত নীতি—"প্রয়োজনের অতিরিক্ত সত্ত্বা অনুমান কোরো না" (Occam's Razor)—এই সূত্রের মাধ্যমে প্রথমবার গাণিতিক ভিত্তি পায়।

লক্ষ করুন, প্রয়োজনীয় নমুনা সংখ্যা নির্ভর করে $ln|mathcalH|$-এর ওপর। তথ্য তত্ত্বে $ln_2|mathcalH|$ হলো একটি হাইপোথিসিসকে এনকোড করার জন্য প্রয়োজনীয় ন্যূনতম বিট সংখ্যা (মডেলের বিবরণী দৈর্ঘ্য বা Description Length)। সুতরাং:

Shakil Mahmud • Research Monograph

> ***ওকামের উপপাদ্য:*** *যদি কোনো সংক্ষিপ্ত এবং সরল মডেল (কম $ln|mathcal{H}|$) পর্যাপ্ত পরিমাণের ট্রেনিং ডেটায় শূন্য ভুল প্রদর্শন করে, তবে গাণিতিকভাবে নিশ্চিত যে সে নতুন অদেখা তথ্যেও সফলভাবে জেনারেলাইজ করবে।*

সরলতা কেবল বিজ্ঞানের নান্দনিক আকাঙ্ক্ষা নয়—এটি পরিসংখ্যানিক শিখনের এক অকাট্য গাণিতিক শর্ত।

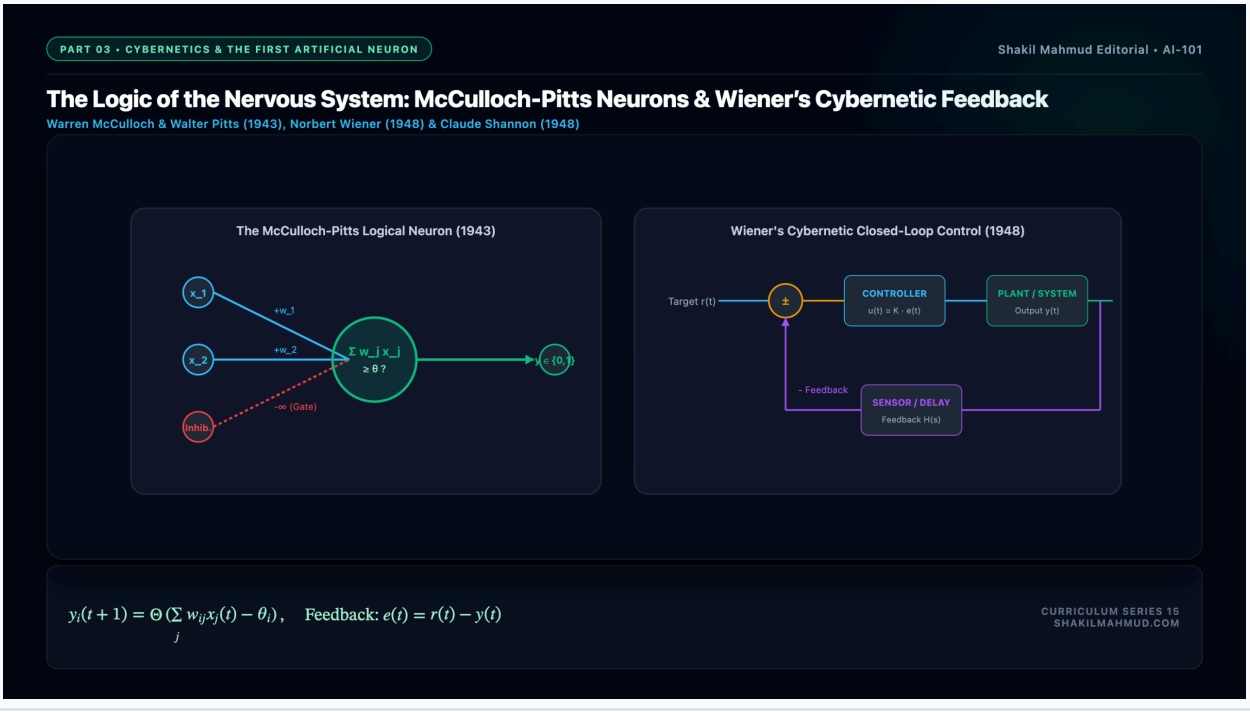
সাইটেশন: মাহমুদ, শাকিল (2026)। "পরিসংখ্যানিক শিখন তত্ত্ব — ভ্যালিয়ান্টের পিএসি ফ্রেমওয়ার্ক, নমুনা জটিলতা এবং ওকামের ক্ষুর"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ২। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-02-statistical-learning-theory-pac-learning-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • তৃতীয় খণ্ড

মডেল সক্ষমতা ও ভিসি ডাইমেনশন — রাডমাখার জটিলতা, গ্রোথ ফাংশন এবং শ্যাটারিং গুণাঙ্ক

মডেল সক্ষমতা ও ভিসি ডাইমেনশন — রাডমাখার জটিলতা, গ্রোথ ফাংশন এবং শ্যাটারিং গুণাঙ্ক

504 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ৩: *ম্যাককুলাক-পিটস বাইনারি লজিক্যাল নিউরন এবং নরবার্ট উইনারের সাইবারনেটিক ক্লোজড-লুপ নেগেটিভ ফিডব্যাক নিয়ন্ত্রণ ব্যবস্থা।*

লেসলি ভ্যালিয়ান্টের সসীম পিএসি মডেল তাত্ত্বিকভাবে অনবদ্য হলেও বাস্তব মেশিন লার্নিং কাজ করে অবিচ্ছিন্ন এবং অসীম হাইপোথিসিস স্পেসে: যেমন $mathbbR^d$ স্পেসের হাইপারপ্লেন, সাপোর্ট ভেক্টর মেশিন বা অসীম প্যারামিটার সমন্বিত নিউরাল নেটওয়ার্ক। অসীম হাইপোথিসিস স্পেসের ক্ষেত্রে $|mathcalH| = infty$, ফলে সসীম সূত্রের $frac1epsilonln|mathcalH|$ অসীমে বিস্ফোরিত হয়। অবিচ্ছিন্ন ফাংশন স্পেসের প্রকৃত প্রকাশক্ষমতা পরিমাপের জন্য সোভিয়েত গণিতবিদ ভ্লাদিমির ভ্যাপনিক এবং আলেক্সি চেরভোনেনকিস উদ্ভাবন করেন **ভিসি ডাইমেনশন তত্ত্ব (VC Dimension Theory)**।

১. শ্যাটারিং (Shattering)-এর ধারণা

ধরা যাক $mathcalH$ হলো একটি বাইনারি ক্লাসিফায়ার ক্লাস $h : mathcalXto-1, +1$। $mathcalX$ স্পেসের m সংখ্যক বিন্দুর একটি সেট $C = x_1, dots, x_m$-এর ওপর $mathcalH$ সম্ভাব্য যতগুলো লেবেল বিন্যাস তৈরি করতে পারে তা হলো:

$$mathcalH_|C = (h(x_1), dots, h(x_m)) : hinmathcalH subseteq-1, +1^m$$

যদি হাইপোথিসিস ক্লাস $mathcalH$ এই m সংখ্যক বিন্দুর সম্ভাব্য সকল 2^m ধরণের ধনাত্মক ও ঋণাত্মক লেবেলিং তৈরি করতে সক্ষম হয়, তবে বলা হয় যে $mathcalH$ ওই বিন্দু সেটটিকে **শ্যাটার (Shatter)** করেছে:

Shakil Mahmud • Research Monograph

$$|mathcalH_{|C}| = 2^m$$

কোনো সেটকে শ্যাটার করার অর্থ হলো—মডেলটির ওই বিন্দুগুলোর ওপর সম্পূর্ণ নিয়ন্ত্রণ রয়েছে; যেকোনো কম্বিনেশন সে জ্যামিতিকভাবে আলাদা করতে পারে।

২. ভিসি ডাইমেনশনের সংজ্ঞা

একটি হাইপোথিসিস ক্লাস $mathcalH$-এর **ভিসি ডাইমেনশন (d_{textVC})** হলো সর্বোচ্চ যতগুলো বিন্দুর সেটকে $mathcalH$ শ্যাটার করতে পারে তার আকার:

$$d_{textVC}(mathcalH) = maxminmathbbN : existsCsubsetmathcalX, |C| = m, |mathcalH_{|C}| = 2^m$$

যদি ক্লাসটি যেকোনো আকারের বিন্দু সেটকে শ্যাটার করতে পারে, তবে তার ভিসি ডাইমেনশন অসীম।

উদাহরণ: $mathbbR^d$-তে রৈখিক হাইপারপ্লেন

দ্বিমাত্রিক তলে ($d = 2$) সরলরেখার মাধ্যমে ক্লাসিফিকেশন বিবেচনা করি: $h(x) = textsign(wcdotx + b)$।

- দ্বিমাত্রিক তলে যেকোনো ৩টি অসমরেখ বিন্দুকে একটি সরলরেখা দিয়ে সম্ভাব্য সকল $2^3 = 8$ উপায়ে পৃথক করা সম্ভব। সুতরাং $d_{textVC}ge3$।
- কিন্তু *র‍্যাডনের উপপাদ্য (Radon's Theorem)* অনুযায়ী, তলের যেকোনো ৪টি বিন্দুকে সরলরেখা দিয়ে শ্যাটার করা অসম্ভব (এক্সওআর সমস্যা—যেখানে বিপরীত কোণায় একই রঙের বিন্দু থাকলে একটি সরলরেখা তা আলাদা করতে পারে না)।

সাধারণভাবে, d-মাত্রিক স্পেসে রৈখিক হাইপারপ্লেনের ভিসি ডাইমেনশন হলো:

$$d_{textVC}(textHyperplanesinmathbbR^d) = d + 1$$

৩. সাউয়ারের লেমা: সূচকীয় থেকে বহুপদী রূপান্তর

গ্রোথ ফাংশন $Pi_{mathcalH}(m)$ পরিমাপ করে m সংখ্যক বিন্দুতে মডেলটি সর্বোচ্চ কতগুলো লেবেল তৈরি করতে পারে। $mled_{textVC}$ পর্যন্ত এটি সূচকীয় গতিতে বৃদ্ধি পায়: $Pi_{mathcalH}(m) = 2^m$। কিন্তু যখনই বিন্দুর সংখ্যা ভিসি সীমা অতিক্রম করে (m > d_{text{VC}}), তখনই এক অবিশ্বাস্য রূপান্তর ঘটে:

> ***সাউয়ার-শেলাহ লেমা (Sauer's Lemma):*** *যদি d_{text{VC}}(mathcal{H}) = d < infty হয়, তবে $mged$-এর জন্য:*
>
> $$Pi_{mathcalH}(m)lesum_{i=0}^{d}binommileft(fracmdright)^d$$

এই সূত্রটি প্রমাণ করে যে ভিসি সীমা অতিক্রম করার পর হাইপোথিসিসের সক্ষমতা হঠাৎ সূচকীয় (2^m) থেকে বহুপদীতে (m^d) নেমে আসে। এই আকস্মিক সীমাবদ্ধতাই মডেলকে অন্ধের মতো নয়েজ মুখস্থ করা থেকে বিরত রেখে সাধারণ জ্ঞান শিখতে বাধ্য করে।

৪. মৌলিক ভিসি জেনারেলাইজেশন বাউন্ড

সাউয়ারের লেমা প্রয়োগ করে ভ্যাপনিক ও চেরভোনেনকিস আধুনিক পরিসংখ্যানিক শিখনের সবচেয়ে গুরুত্বপূর্ণ বাউন্ড প্রমাণ করেন:

$$R(h) le R_{text emp}(h) + sqrt frac 8 d_{text VC} ln(2en/d_{text VC}) + 8ln(4/delta) n$$

এই সমীকরণটি দেখায় যে নতুন তথ্যে মডেলের আসল ভুল ট্রেনিং ভুলের সাথে একটি ক্যাপাসিটি পেনাল্টির যোগফল দ্বারা নির্ধারিত হয়, যা $mathcal O left(sqrt frac d_{text VC} n right)$ হারে হ্রাস পায়। সবচেয়ে বিস্ময়কর বিষয় হলো—মডেল কতটা ভালো জেনারেলাইজ করবে তা ইনপুট স্পেসের ডাইমেনশনের ওপর নয়, বরং মডেলের নিজস্ব ভিসি ডাইমেনশনের ওপর নির্ভর করে।

❦

সাইটেশন: মাহমুদ, শাকিল (2026)। "মডেল সক্ষমতা ও ভিসি ডাইমেনশন — রাডমাখার জটিলতা, গ্রোথ ফাংশন এবং শ্যাটারিং গুণাঙ্ক"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ৩। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-03-model-capacity-vc-dimension-bn/

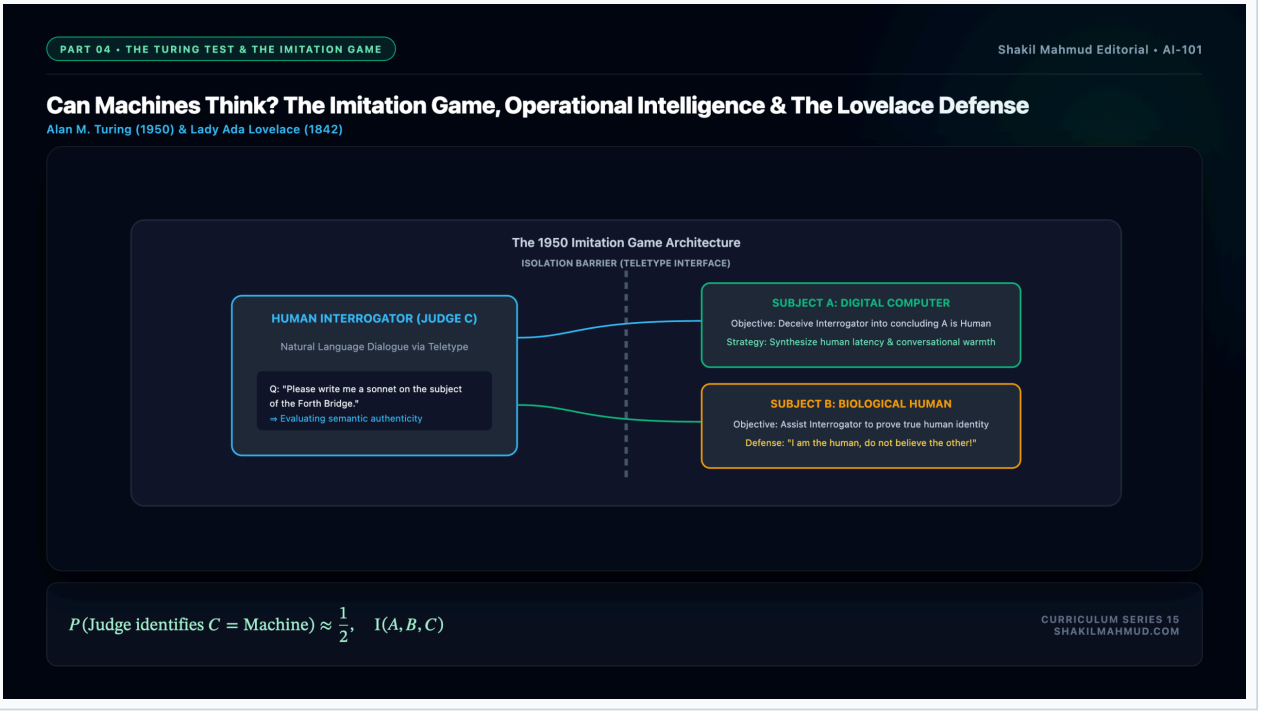
Shakil Mahmud • Research Monograph

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • চতুর্থ খণ্ড

বায়াস-ভ্যারিয়েন্স ভারসাম্য — প্রত্যাশিত পূর্বাভাস ত্রুটির বিশ্লেষণ ও ওভারফিটিং গতিবিজ্ঞান

বায়াস-ভ্যারিয়েন্স ভারসাম্য — প্রত্যাশিত পূর্বাভাস ত্রুটির বিশ্লেষণ ও ওভারফিটিং গতিবিজ্ঞান

418 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ৪: *অ্যালান টুরিংয়ের ১৯৫০ সালের ইমিটেশন গেম ট্রাইপার্টাইট স্থাপত্য, যা ব্লাইন্ড টেলিটাইপ কথোপকথনের মাধ্যমে কৃত্রিম বুদ্ধিমত্তা মূল্যায়ন করে।*

মেশিন লার্নিংয়ের যেকোনো গবেষক বা প্রকৌশলীকে একটি মৌলিক টানাপোড়েনের মধ্য দিয়ে যেতে হয়—মডেলটি কি এত সরল যে আসল সত্য দেখতে পাচ্ছে না, নাকি এত জটিল যে সে র‍্যান্ডম নয়েজের মধ্যেও মিথ্যা প্যাটার্ন আবিষ্কার করছে? ধ্রুপদী পরিসংখ্যানিক শিখন তত্ত্বে এই টানাপোড়েনকে বলা হয় **বায়াস-ভ্যারিয়েন্স ভারসাম্য (Bias-Variance Tradeoff)**: পূর্বাভাসের গড় দ্বিঘাত ত্রুটিকে দুটি প্রতিদ্বন্দ্বী শক্তি এবং একটি অপূরণীয় নয়েজ ফ্লোরে বিভক্ত করার অকাট্য গাণিতিক সমীকরণ।

১. ডেটা উৎপাদন ও এস্টিমেটর

ধরা যাক আমাদের লক্ষ্য ভেরিয়েবল $yinmathbbR$ একটি অন্তর্নিহিত নিখুঁত ফাংশন $f^*(x)$ এবং শূন্য-গড় বিশিষ্ট র‍্যান্ডম নয়েজ $epsilon$-এর যোগফল দ্বারা তৈরি:

$$y = f^*(x) + epsilon, quadtextযেখানে quadmathbbE[epsilon] = 0, quadtextVar(epsilon) = sigma^2, quadepsilonperpx$$

আমাদের শিখন অ্যালগরিদম একটি ট্রেনিং সেট $SimmathcalD^n$ দেখে একটি অনুমিত ফাংশন $hatf_S(x)$ তৈরি করে। যেহেতু ট্রেনিং সেট S একটি দৈব নমুনা, তাই মডেল $hatf_S(x)$ স্বয়ং একটি **দৈব চলক (Random Variable)**।

সম্ভাব্য সকল ভিন্ন ভিন্ন ট্রেনিং সেটের ওপর মডেলটির গড় অনুমানকে সংজ্ঞায়িত করা হয়:

$$bar f(x) = mathbb E_S[hat f_S(x)]$$

২. প্রত্যাশিত পূর্বাভাস ত্রুটির গাণিতিক প্রতিপাদন

যেকোনো টেস্ট পয়েন্ট x_0-তে প্রত্যাশিত দ্বিঘাত ত্রুটি (Expected Prediction Error) বিশ্লেষণ করা যাক:

$$text{EPE}(x_0) = mathbb{E}_{S, epsilon}left[(y – hat{f}_S(x_0))^2 right]$$

আমরা $y = f^*(x_0) + epsilon$ বসিয়ে এবং বন্ধনীর ভেতরে $bar f(x_0)$ যোগ ও বিয়োগ করে পাই:

$$y – hat{f}_S(x_0) = (f^*(x_0) – bar{f}(x_0)) + (bar{f}(x_0) – hat{f}_S(x_0)) + epsilon$$

বর্গ সম্প্রসারণ এবং প্রত্যাশা (Expectation) গ্রহণ করার পর, যেহেতু নয়েজ $epsilon$-এর গড় শূন্য এবং এটি S থেকে সম্পূর্ণ স্বাধীন, তাই সকল ক্রস-টার্ম বা গুণফল শূন্য হয়ে যায়:

$$text{EPE}(x_0) = underbrace{(f^*(x_0) – bar{f}(x_0))^2}_{text{বায়াস}^২} + underbrace{mathbb{E}_Sleft[(hat{f}_S(x_0) – bar{f}(x_0))^2right]}_{text{ভ্যারিয়েন্স}} + underbrace{sigma^2}_{text{অপূরণীয় নয়েজ}}$$

৩. তিনটি উপাদানের শারীরস্থান

- **বায়াস ($textBias$):** মডেলের কাঠামোগত দুর্বলতার কারণে সৃষ্ট ভ্রান্তি। অসীম সংখ্যক ভিন্ন ডেটাসেটে মডেলটির গড় পূর্বাভাস এবং আসল সত্য $f^*(x)$-এর মধ্যকার ব্যবধানই হলো বায়াস। বায়াস বেশি হওয়া মানে মডেলটি বাস্তবতাকে ধরতে পারছে না (**আন্ডারফিটিং**)।
- **ভ্যারিয়েন্স ($textVariance$):** নির্দিষ্ট নমুনা সেটের ওপর মডেলটির সংবেদনশীলতা। ভিন্ন ভিন্ন ট্রেনিং সেট দেখলে মডেলটি কতটা এলোমেলোভাবে পরিবর্তিত হয় এটি তার পরিমাপ। ভ্যারিয়েন্স বেশি হওয়া মানে মডেলটি ডেটার নয়েজকেও অন্ধের মতো বিশ্বাস করে ফেলেছে (**ওভারফিটিং**)।
- **অপূরণীয় নয়েজ ($sigma^2$):** বাস্তব জগতের সহজাত বিশৃঙ্খলা। পৃথিবীর কোনো অ্যালগরিদমের পক্ষেই এই $sigma^2$ সীমার নিচে ত্রুটি নামিয়ে আনা সম্ভব নয়।

৪. ধ্রুপদী ভারসাম্য বনাম আধুনিক ডাবল ডিসেন্ট

ধ্রুপদী পরিসংখ্যানে মডেলের জটিলতা বৃদ্ধির সাথে সাথে:

$$textজটিলতা uparrow implies text বায়াস^২ downarrow, quad text ভ্যারিয়েন্স uparrow$$

ফলে মোট টেস্ট ভুলের গ্রাফটি ইংরেজি “U” অক্ষরের মতো রূপ নেয়, যার সর্বনিম্ন বিন্দুটিই হলো সর্বোত্তম জটিলতা। তবে গভীর নিউরাল নেটওয়ার্কে দেখা যায় যে মডেল যখন ট্রেনিং ডেটা শতভাগ নিখুঁত মুখস্থ করার সীমা ছাড়িয়ে যায় (ইন্টারপোলেশন থ্রেশহোল্ড), তখন গ্রাফটি দ্বিতীয়বার নামতে শুরু করে—যাকে আধুনিক এআই তত্ত্বে বলা হয় **ডাবল ডিসেন্ট (Double Descent)**।

সাইটেশন: মাহমুদ, শাকিল (2026)। "বায়াস-ভ্যারিয়েন্স ভারসাম্য — প্রত্যাশিত পূর্বাভাস ত্রুটির বিশ্লেষণ ও ওভারফিটিং গতিবিজ্ঞান"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ৪। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-04-bias-variance-tradeoff-loss-geometry-bn/

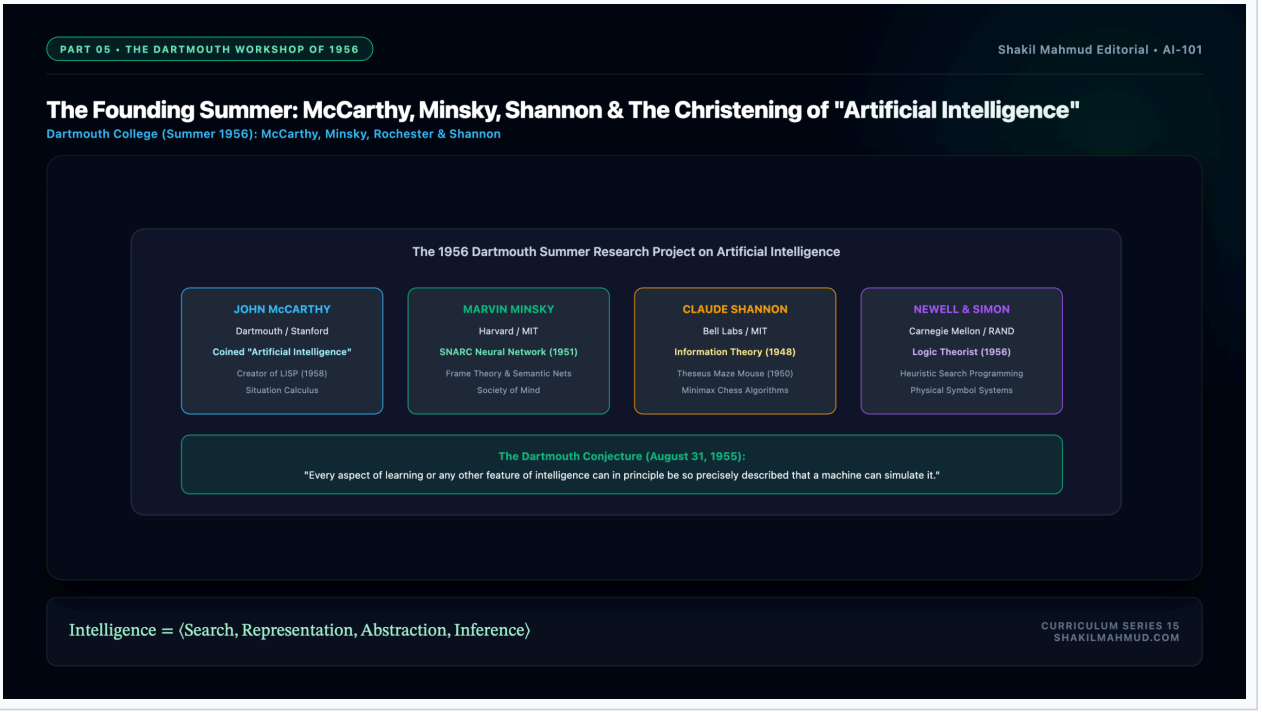
Shakil Mahmud • Research Monograph

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • পঞ্চম খণ্ড

বহু-মাত্রিক উত্তল অপ্টিমাইজেশন — গ্রেডিয়েন্ট ডিসেন্ট, সাব-গ্রেডিয়েন্ট এবং রূপান্তর গতিবিজ্ঞান

বহু-মাত্রিক উত্তল অপ্টিমাইজেশন — গ্রেডিয়েন্ট ডিসেন্ট, সাব-গ্রেডিয়েন্ট এবং রূপান্তর গতিবিজ্ঞান

502 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ৫: *১৯৫৬ সালের ডার্টমাউথ এআই সম্মেলন: জন ম্যাককার্থি, মার্ভিন মিনস্কি, ক্লড শ্যানন এবং নিউয়েল-সাইমনের যুগান্তকারী প্রতিষ্ঠা।*

লিনিয়ার রিগ্রেশন ও লজিস্টিক মডেল থেকে সাপোর্ট ভেক্টর মেশিন পর্যন্ত প্রায় প্রতিটি ক্লাসিক্যাল মেশিন লার্নিং সমস্যা মূলত একটি অপ্টিমাইজেশন বা অবমূল্যায়ন সমস্যা। যখন লক্ষ্য লস ফাংশনটি উত্তল বা কনভেক্স (Convex) হয়, তখন স্থানীয় ফাঁদ (Local Minima), স্যাডল পয়েন্ট এবং বিভ্রান্তিকর মালভূমির জটিলতা উবে যায়। বহু-মাত্রিক উত্তল জ্যামিতিতে যেকোনো স্থানীয় ন্যূনতম মানই হলো পরম বৈশ্বিক ন্যূনতম মান (Global Minimum), এবং অ্যালগরিদমের গতি নির্ধারিত হয় কঠোর গাণিতিক নিয়মে।

১. উত্তল সেট ও ফাংশনের জ্যামিতি

একটি সেট $mathcal C subseteq mathbb R^d$-কে **উত্তল (Convex)** বলা হয় যদি সেটের যেকোনো দুটি বিন্দু $x, y in mathcal C$-এর সংযোগকারী সরলরেখাংশ সম্পূর্ণভাবে সেটের ভেতরে অবস্থান করে:

theta x + (1 – theta) y in mathcal{C} quad forall theta in [0, 1]

একটি ফাংশন $f(x)$ উত্তল হবে যদি তার এপিগ্রাফ (ফাংশনের গ্রাফের ওপরের অংশ) একটি উত্তল সেট হয়। ডিফারেনশিয়েবল ফাংশনের ক্ষেত্রে এর অর্থ হলো—ফাংশনের যেকোনো বিন্দুতে অঙ্কিত স্পর্শক বা ফার্স্ট-অর্ডার টেলর সিরিজ সর্বদা মূল ফাংশনের নিচে অবস্থান করে:

f(y) ge f(x) + nabla f(x)^T (y – x)

২. লিপশিৎস মসৃণতা ও স্ট্রং কনভেক্সিটি

অপ্টিমাইজেশন অ্যালগরিদম কত দ্রুত লক্ষ্যে পৌঁছাবে তা পরিমাপ করতে ফাংশনের বক্রতাকে দুটি মানদণ্ডে বিভক্ত করা হয়:

1. **L-লিপশিৎস মসৃণতা (Lipschitz Smoothness):** গ্রেডিয়েন্টের পরিবর্তনের গতি সীমাবদ্ধ। এর ফলে ফাংশনটি যেকোনো বিন্দুতে একটি নির্দিষ্ট প্যারাবোলয়েড বা দ্বিঘাত বাটির (L) নিচে আবদ্ধ থাকে:

$$f(y) le f(x) + nabla f(x)^T (y – x) + frac{L}{2} |y – x|^2$$

2. **mu-স্ট্রং কনভেক্সিটি (Strong Convexity):** ফাংশনটি কেবল উত্তল নয়, বরং এটি কমপক্ষে mu বক্রতা বিশিষ্ট একটি দ্বিঘাত বাটির চেয়ে দ্রুত বৃদ্ধি পায়:

$$f(y) ge f(x) + nabla f(x)^T (y – x) + frac{mu}{2} |y – x|^2$$

এটি প্রমাণ করে যে ফাংশনের একটিই অনন্য গ্লোবাল মিনিমাম w^* আছে এবং মিনিমামের বাইরে গ্রেডিয়েন্ট কখনোই শূন্য হতে পারে না।

৩. গ্রেডিয়েন্ট ডিসেন্টের রূপান্তর সূত্র

স্ট্যান্ডার্ড গ্রেডিয়েন্ট ডিসেন্ট প্রতি ধাপে লার্নিং রেট eta অনুযায়ী নেগেটিভ গ্রেডিয়েন্টের দিকে অগ্রসর হয়:

$$w_{t+1} = w_t – eta nabla f(w_t)$$

গাণিতিক বিশ্লেষণ থেকে দুটি ভিন্ন গতির রূপান্তর পাওয়া যায়:

- **সাধারণ উত্তল ও মসৃণ ফাংশনে:** ত্রুটি হ্রাস পায় সাব-লিনিয়ার $mathcalO(1/T)$ হারে:

$$f(w_T) – f(w^*) le frac{L |w_0 – w^*|^2}{2T}$$

- **স্ট্রং কনভেক্স ও মসৃণ ফাংশনে:** দূরত্ব হ্রাস পায় জ্যামিতিক প্রগমনে দ্রুত গতিতে (Linear Convergence):

$$|w_T – w^*|^2 le left(frac{kappa – 1}{kappa + 1} right)^{2T} |w_0 – w^*|^2$$

৪. কন্ডিশন নম্বর $kappa$ ও আঁকাবাঁকা গতি

হেসিয়ান ম্যাট্রিক্সের বৃহত্তম ও ক্ষুদ্রতম আইগেনভ্যালুর অনুপাত হলো **কন্ডিশন নম্বর ($kappa = L/mu$)**।

- যখন $kappaapproxx1$, তখন লেভেল কার্ভগুলো বৃত্তাকার হয় এবং গ্রেডিয়েন্ট সরাসরি কেন্দ্রের দিকে নির্দেশ করে, ফলে রূপান্তর হয় চোখের পলকে।
- যখন $kappaggx1$, তখন লেভেল কার্ভগুলো সরু দীর্ঘ উপবৃত্তাকার গিরিখাত তৈরি করে। গ্রেডিয়েন্ট সর্বনিম্ন বিন্দুর দিকে না গিয়ে খাতের দুই দেয়ালে এলোমেলোভাবে আছড়ে পড়ে (Zig-zagging), যার ফলে অ্যালগরিদমের গতি অত্যন্ত ধীর হয়ে যায়।

৫. সাব-গ্রেডিয়েন্ট ক্যালকুলাস

ল্যাসোর L_1 রেগুলারাইজেশন বা এসভিএম-এর হিঞ্জ লসের মতো ফাংশনগুলো কোণায় ডিফারেনশিয়েবল নয়। এই সমস্যা সমাধানে ব্যবহৃত হয় **সাব-গ্রেডিয়েন্ট (Subgradient)** ধারণা:

$$partial f(x) = {g in mathbb{R}^d : f(y) ge f(x) + g^T (y – x) quad forall y}$$

Shakil Mahmud • Research Monograph

কোণার বিন্দুতে একটি স্পর্শকের বদলে সম্ভাব্য বহু স্পর্শকের একটি সেট তৈরি হয়। যেকোনো একটি বৈধ সাব-গ্রেডিয়েন্ট g_t অনুসরণ করে সাব-গ্রেডিয়েন্ট ডিসেন্ট $mathcalO(1/sqrtT)$ হারে সফলভাবে বৈশ্বিক অনুকূল মানে পৌঁছাতে সক্ষম হয়।



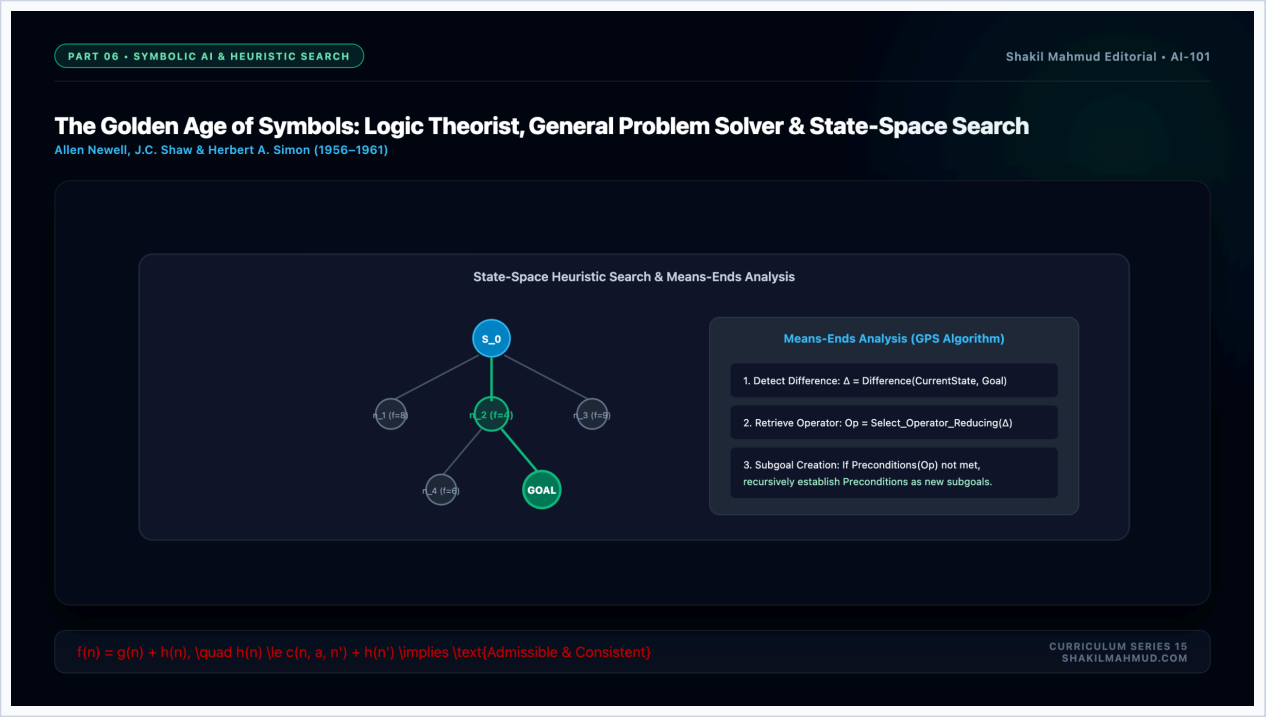
সাইটেশন: মাহমুদ, শাকিল (2026)। "বহু-মাত্রিক উত্তল অপ্টিমাইজেশন — গ্রেডিয়েন্ট ডিসেন্ট, সাব-গ্রেডিয়েন্ট এবং রূপান্তর গতিবিজ্ঞান"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ৫। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-05-convex-optimization-gradient-descent-convergence-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • ষষ্ঠ খণ্ড

স্টোকাস্টিক অপ্টিমাইজেশন — মিনি-ব্যাচ এসজিডি, মোমেন্টাম এবং অ্যাডাপ্টিভ মোমেন্টস (অ্যাডাম)

স্টোকাস্টিক অপ্টিমাইজেশন — মিনি-ব্যাচ এসজিডি, মোমেন্টাম এবং অ্যাডাপ্টিভ মোমেন্টস (অ্যাডাম)

420 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ৬: *স্টেট-স্পেস হিউরিস্টিক সার্চ ট্রি এবং জেনারেল প্রবলেম সলভারের (GPS) মিন্স-এন্ডস অ্যানালাইসিস অ্যালগরিদম।*

বিশাল আকারের আধুনিক ডেটাসেটে যেখানে কোটি কোটি ডেটা পয়েন্ট থাকে, সেখানে প্রতি ধাপে পুরো ডেটাসেটের ওপর নিখুঁত গ্রেডিয়েন্ট গণনা করা কম্পিউটেশনালি অসম্ভব। মেশিন লার্নিং এই চ্যালেঞ্জ জয় করেছে দৈবতার হাত ধরে: **স্টোকাস্টিক গ্রেডিয়েন্ট ডিসেন্ট (SGD)** এবং এর আধুনিক রূপান্তর। পুরো ডেটাসেটের বদলে একটি বা কয়েকটি এলোমেলো নমুনা নিয়ে কাজ করে এসজিডি প্রমাণ করেছে যে র‍্যান্ডম নয়েজ অপ্টিমাইজেশনের শত্রু নয়, বরং দ্রুততম বন্ধু।

১. স্টোকাস্টিক গ্রেডিয়েন্টের অনুমান

আমাদের মূল লক্ষ্য হলো সমগ্র ডেটাসেটের গড় লস কমানো:

$$f(w) = frac1nsum_{i=1}^{n} f_i(w)$$

পুরো n টি ডেটা না দেখে এসজিডি প্রতি ধাপে যেকোনো একটি ডেটা পয়েন্ট i_t বেছে নেয় এবং কেবল তারই গ্রেডিয়েন্ট দিয়ে মডেল আপডেট করে:

w_{t+1} = w_t – eta_t nabla f_{i_t}(w_t)

Shakil Mahmud • Research Monograph

সবচেয়ে গুরুত্বপূর্ণ বিষয় হলো—এই স্টোকাস্টিক গ্রেডিয়েন্ট একটি **আনবায়াসড এস্টিমেটর (Unbiased Estimator)**, অর্থাৎ এর প্রত্যাশিত গড় মান আসল গ্রেডিয়েন্টের সমান:

$$mathbbE[g_t] = nablaf(w_t)$$

তবে একক ডেটার ওপর নির্ভর করায় এতে অনেক নয়েজ থাকে। তাই ব্যবহারিক ক্ষেত্রে B আকারের মিনি-ব্যাচ (Mini-batch) ব্যবহার করা হয়, যা ভ্যারিয়েন্সকে B গুণ কমিয়ে আনে।

২. পলিক মোমেন্টাম ও নেস্টেরভ অ্যাক্সিলারেশন

সরু উপত্যকার দেয়ালে অযথা কাঁপাকাঁপি বন্ধ করতে এবং সমতল রাস্তায় গতি বাড়াতে বরিস পলিক (১৯৬৪) নিয়ে আসেন **হেভি-বল মোমেন্টাম (Heavy-Ball Momentum)**, যা পদার্থবিদ্যার ঘর্ষণ ও ভরের ধারণায় পরিচালিত:

v_t = beta v_{t-1} + g_t, quad w_{t+1} = w_t – eta v_t

যেসব দিকে গ্রেডিয়েন্ট দিক পরিবর্তন করে কাঁপছে, সেগুলো একে অপরকে বিয়োগ করে বাতিল করে দেয়; আর যেদিকে নিয়মিত ঢাল বিদ্যমান, সেদিকে বেগ বহুগুণ বৃদ্ধি পায়।

ইউরি নেস্টেরভ (১৯৮৩) এতে যোগ করেন **নেস্টেরভ লুক-অ্যাহেড (NAG)**। বর্তমান অবস্থানে গ্রেডিয়েন্ট না মেপে মোমেন্টামের দিকে একটু এগিয়ে গিয়ে গ্রেডিয়েন্ট পরিমাপ করা হয়:

g_t = nabla f(w_t – eta beta v_{t-1})

এটি উপত্যকার বিপরীত দেয়ালে ওঠার আগেই ব্রেক কষে, যা উত্তল অপ্টিমাইজেশনের গতিকে তাত্ত্বিক সর্বোচ্চ সীমা $mathcalO(1/T^2)$-এ উন্নীত করে।

৩. অ্যাডাপ্টিভ মোমেন্ট এস্টিমেশন: অ্যাডাম (Adam)

উচ্চ মাত্রার স্পেসে প্রতিটি প্যারামিটারের সংবেদনশীলতা ভিন্ন ভিন্ন হয়। একটি একক লার্নিং রেট সবার জন্য প্রযোজ্য হতে পারে না। ২০১৪ সালে ডিডেরিক কিংমা এবং জিমি বা তৈরি করেন **অ্যাডাম (Adam)** অ্যালগরিদম:

1. **প্রথম মোমেন্ট (গতিবেগ):**

m_t = beta_1 m_{t-1} + (1 – beta_1) g_t

2. **দ্বিতীয় মোমেন্ট (স্কেল/বক্রতা):**

v_t = beta_2 v_{t-1} + (1 – beta_2) g_t^2

3. **বায়াস সংশোধন:** যেহেতু $m_0 = 0$ ও $v_0 = 0$ দিয়ে শুরু হয়, তাই প্রথমদিকে মানগুলো কৃত্রিমভাবে ছোট থাকে। এটি ঠিক করতে ভাগ করা হয়:

hat{m}_t = frac{m_t}{1 – beta_1^t}, quad hat{v}_t = frac{v_t}{1 – beta_2^t}

4. **প্যারামিটার আপডেট:**

w_{t+1} = w_t – frac{eta}{sqrt{hat{v}_t} + epsilon} odot hat{m}_t

অ্যাডাম প্রতিটি কোঅর্ডিনেটকে তার নিজস্ব ভ্যারিয়েন্স অনুযায়ী স্বয়ংক্রিয়ভাবে নিয়ন্ত্রণ করে—যে প্যারামিটারে নয়েজ বেশি তার গতি কমিয়ে দেয়, আর যেটিতে কদাচিৎ সিগন্যাল আসে তার গুরুত্ব বাড়িয়ে দেয়। আজ মেশিন লার্নিংয়ের সর্বাধিক জনপ্রিয় অপ্টিমাইজার হলো অ্যাডাম।

সাইটেশন: মাহমুদ, শাকিল (2026)। "স্টোকাস্টিক অপ্টিমাইজেশন — মিনি-ব্যাচ এসজিডি, মোমেন্টাম এবং অ্যাডাপ্টিভ মোমেন্টস (অ্যাডাম)"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ৬। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-06-stochastic-optimization-dynamics-sgd-adam-bn/

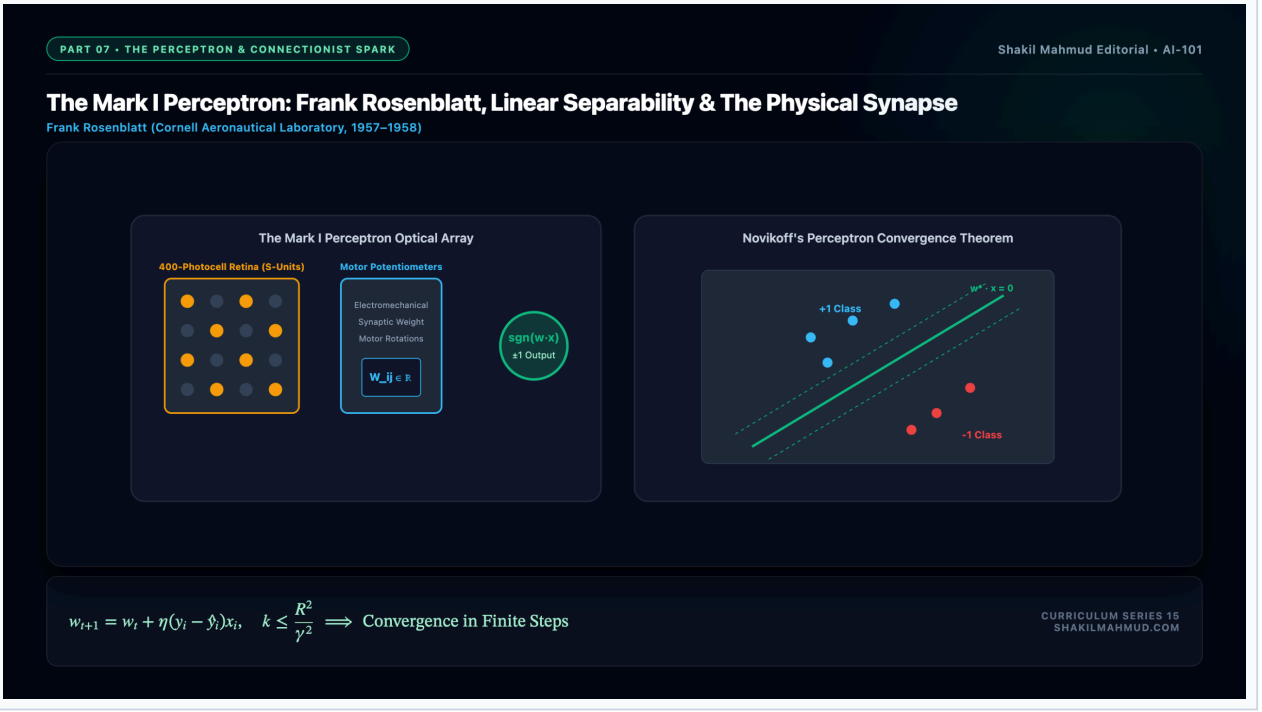
Shakil Mahmud • Research Monograph

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • সপ্তম খণ্ড

রৈখিক বিভাজক ও সার্বজনীন রৈখিক মডেল — ফিশারের এলডিএ এবং এক্সপোনেনশিয়াল ফ্যামিলি

রৈখিক বিভাজক ও সার্বজনীন রৈখিক মডেল — ফিশারের এলডিএ এবং এক্সপোনেনশিয়াল ফ্যামিলি

412 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ৭: *ফ্র্যাঙ্ক রোজেনব্ল্যাটের মার্ক ১ পারসেপট্রন রেটিনা, মোটর পটেনশিওমিটার ওজন এবং নোভিকফের কনভার্জেন্স হাইপারপ্লেন।*

রৈখিক মডেল হলো পরিসংখ্যানিক প্যাটার্ন চেনার প্রথম প্রামাণ্য ভিত্তিপ্রস্তর। আধুনিক ডিপ নিউরাল নেটওয়ার্কের বহু পূর্বে স্যার রোনাল্ড ফিশার, হ্যারল্ড হোটেলিং এবং নেল্ডার-ওয়েডারবার্ন রৈখিক প্রক্ষেপণ (Linear Projection) এবং সম্ভাব্যতা ঘনত্বের মধ্যে এক অবিচ্ছেদ্য গাণিতিক সম্পর্ক তৈরি করেছিলেন। **ফিশারের লিনিয়ার ডিসক্রিমিন্যান্ট অ্যানালাইসিস (LDA)** এবং **জেনারেলাইজড লিনিয়ার মডেল (GLM)**-এর মাধ্যমে আমরা অনুধাবন করি কীভাবে জ্যামিতি এবং পরিসংখ্যান একত্রিত হয়।

১. ফিশারের লিনিয়ার ডিসক্রিমিন্যান্ট: স্ক্যাটার অনুপাত সর্বোচ্চকরণ

১৯৩৬ সালে রোনাল্ড ফিশার ক্লাসিফিকেশনকে একটি অপ্টিমাল ডাইমেনশন হ্রাসের সমস্যা হিসেবে দেখেন: এমন একটি ভেক্টর $winmathbbR^d$ খুঁজে বের করতে হবে যা ডেটাকে এমনভাবে প্রক্ষেপণ করবে যাতে দুই ক্লাসের কেন্দ্রের দূরত্ব সর্বোচ্চ হয় এবং একই সাথে প্রতিটি ক্লাসের ভেতরের বিস্তৃতি (Scatter) সর্বনিম্ন হয়।

দুই ক্লাসের কেন্দ্রের মধ্যবর্তী দূরত্বের ম্যাট্রিক্স হলো **বিটুইন-ক্লাস স্ক্যাটার ম্যাট্রিক্স** (S_B):

S_B = (mu_1 – mu_2)(mu_1 – mu_2)^T

আর ক্লাসের ভেতরের সামগ্রিক ভ্যারিয়েন্স হলো **উইদিন-ক্লাস স্ক্যাটার ম্যাট্রিক্স** (S_W):

S_W = sum_{x in mathcal{C}_1} (x – mu_1)(x – mu_1)^T + sum_{x in mathcal{C}_2} (x – mu_2)(x – mu_2)^T

ফিশার এর অনুপাত সর্বোচ্চ করার সমীকরণ লেখেন (Rayleigh Quotient):

$$J(w) = fracw^T S_B w w^T S_W w$$

একে ব্যবকলন করে শূন্য ধরলে এক বিস্ময়কর সরল সমীকরণ বেরিয়ে আসে:

w^* propto S_W^{-1} (mu_1 – mu_2)

এটিই হলো ফিশারের বিখ্যাত সূত্র—যা উইদিন-ক্লাস কোভ্যারিয়েন্সের ইনভার্স দিয়ে দুই গড়ের দূরত্বকে গুণ করে পাওয়া যায়।

২. জেনারেটিভ বনাম ডিসক্রিমিনেটিভ মডেল

ক্লাসিফায়ার মূলত দুটি দর্শনে বিভক্ত:

- **জেনারেটিভ ক্লাসিফায়ার (যেমন Gaussian LDA):** পুরো জয়েন্ট ডিস্ট্রিবিউশন $P(x, y)$ নির্ণয় করে। যদি ধরে নেওয়া হয় উভয় ক্লাস সমান কোভ্যারিয়েন্স $Sigma$ বিশিষ্ট নরমাল ডিস্ট্রিবিউশন মেনে চলে, তবে পোস্টেরিয়র লজিক ঠিক ফিশারের সমীকরণেই ফিরে আসে: w = Sigma^{-1}(mu_1 – mu_0)।
- **ডিসক্রিমিনেটিভ ক্লাসিফায়ার (যেমন Logistic Regression):** ডেটা কীভাবে তৈরি হলো তা নিয়ে মাথা না ঘামিয়ে সরাসরি শর্তাধীন সম্ভাব্যতা $P(ymidx)$ নির্ণয় করে, যা ডেটার ওপর কোনো অপ্রয়োজনীয় নরমাল অনুমান চাপায় না।

৩. এক্সপোনেনশিয়াল ফ্যামিলি ও জিএলএম (GLM)

১৯৭২ সালে নেল্ডার এবং ওয়েডারবার্ন রিগ্রেশন, লজিস্টিক ক্লাসিফিকেশন এবং পয়সন মডেলকে **জেনারেলাইজড লিনিয়ার মডেল (GLM)**-এর অধীনে একীভূত করেন। একটি ডিস্ট্রিবিউশনকে এক্সপোনেনশিয়াল ফ্যামিলির অংশ বলা হয় যদি তার রূপ হয়:

p(y; eta) = b(y) expleft(eta^T T(y) – A(eta) right)

এখানে eta হলো ন্যাচারাল প্যারামিটার এবং $A(eta)$ হলো লগ-পার্টিশন ফাংশন। $A(eta)$-এর প্রথম ডেরিভেটিভ হলো ডিস্ট্রিবিউশনের গড় ($mathbbE[y]$) এবং দ্বিতীয় ডেরিভেটিভ হলো ভ্যারিয়েন্স।

জিএলএম-এ একটি **লিংক ফাংশন** $g(mu)$ রৈখিক পূর্বাভাস $eta = w^T x$-এর সাথে গড় মানকে যুক্ত করে:

$$eta = g(mu) iff mu = g^{-1}(w^T x)$$

লিনিয়ার রিগ্রেশনে এটি আইডেন্টিটি ফাংশন, লজিস্টিক রিগ্রেশনে এটি সিগময়েড/লজিট ফাংশন, এবং কাউন্ট মডেলে এটি লগ ফাংশন। এই একক রূপরেখা প্রতিটি মডেলের অবজেক্টিভ ফাংশনকে কঠোরভাবে উত্তল (Strictly Concave) রাখে, যার ফলে নিউটন-র‍্যাফসন পদ্ধতি নিশ্চিতভাবে সঠিক সমাধানে পৌঁছাতে পারে।

সাইটেশন: মাহমুদ, শাকিল (2026)। "রৈখিক বিভাজক ও সার্বজনীন রৈখিক মডেল — ফিশারের এলডিএ এবং এক্সপোনেনশিয়াল ফ্যামিলি"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ৭। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-07-linear-discriminants-glms-exponential-families-bn/

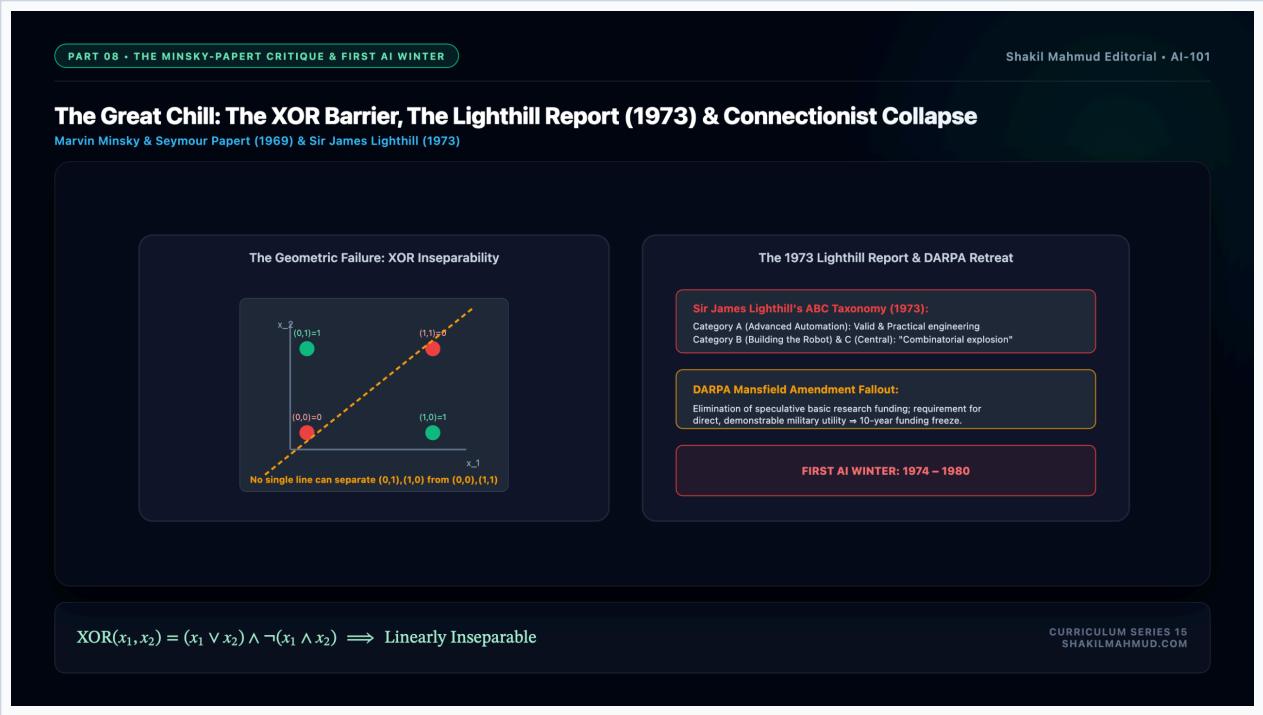
Shakil Mahmud • Research Monograph

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • অষ্টম খণ্ড

মার্জিনের জ্যামিতি — সাপোর্ট ভেক্টর মেশিন, কেকেটি শর্তাবলি এবং সাপোর্ট ভেক্টর

মার্জিনের জ্যামিতি — সাপোর্ট ভেক্টর মেশিন, কেকেটি শর্তাবলি এবং সাপোর্ট ভেক্টর

529 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ৮: *মিনস্কি ও প্যাপার্টের প্রমাণিত এক্সওআর (XOR) লিনিয়ার ইনসেপারেবিলিটি এবং ১৯৭৩ সালের লাইটহিল রিপোর্ট পরবর্তী প্রথম এআই শীত।*

দুটি পৃথক ক্লাসের মাঝখান দিয়ে আঁকা যায় এমন অসীম সংখ্যক সরলরেখার মধ্যে কোনটি গাণিতিকভাবে সর্বশ্রেষ্ঠ? ফ্র্যাঙ্ক রোজেনব্ল্যাটের পারসেপ্ট্রন যে সরলরেখা খুঁজে পেত, তা প্রায়শই ডেটা পয়েন্টের খুব কাছাকাছি চলে আসত—ফলে সামান্য নয়েজ পেলেই মডেলটি ভুল করত। ১৯৯০-এর দশকে ভ্লাদিমির ভ্যাপনিক, বের্নহার্ড বোসার এবং কোরিনা কর্টেস নিয়ে আসেন **সাপোর্ট ভেক্টর মেশিন (SVM)**: এক অনবদ্য জ্যামিতিক কাঠামো যা দুই ক্লাসের মাঝখানের দূরত্ব বা মার্জিন (Margin) সর্বোচ্চ করে উত্তল অপ্টিমাইজেশনের মাধ্যমে বিশ্বস্ততম সমাধান নিশ্চিত করে।

১. জ্যামিতিক মার্জিনের ধারণা

ধরা যাক ট্রেনিং ডেটা (x_i, y_i) রৈখিকভাবে বিভাজ্য, যেখানে $y_i in -1, +1$। একটি হাইপারপ্লেন সংজ্ঞায়িত হয় $wcdotx + b = 0$ দ্বারা। যেকোনো ডেটা পয়েন্ট থেকে এই হাইপারপ্লেনের লম্ব ইউক্লিডীয় দূরত্ব হলো:

$$gamma_i = fracy_i(wcdotx_i + b)|w|$$

সবচেয়ে কাছের ডেটা পয়েন্ট থেকে হাইপারপ্লেনের দূরত্বই হলো **জ্যামিতিক মার্জিন ($gamma$)**। ভ্যাপনিক এমন একটি হাইপারপ্লেন খুঁজলেন যা এই ন্যূনতম দূরত্বকে সর্বোচ্চ করে:

$$max_{w,b} min_i frac y_i(wcdot x_i + b)|w|$$

ক্যানোনিকাল স্কেলিং অনুযায়ী সবচেয়ে কাছের বিন্দুর জন্য $y_i(wcdot x_i + b) = 1$ ধরে নিলে মার্জিনের মান দাঁড়ায় $frac 1|w|$। সুতরাং মার্জিন সর্বোচ্চ করা মানে হলো $frac 12|w|^2$-কে সর্বনিম্ন করা।

২. প্রাইমাল উত্তল কোয়াড্রেটিক সমস্যা

এখান থেকেই তৈরি হয় এসভিএম-এর মূল অবজেক্টিভ ফাংশন:

$$min_{w,b} frac 12|w|^2 quad text শর্ত : quad y_i(wcdot x_i + b) ge 1 quad forall i$$

এই সমীকরণটির গাণিতিক সৌন্দর্য অপরিসীম: লক্ষ্য ফাংশনটি কঠোরভাবে উত্তল এবং শর্তগুলো রৈখিক। নিউরাল নেটওয়ার্কের মতো এতে কোনো লোকাল ট্র্যাপ বা স্যাডল পয়েন্ট থাকে না—এর একটিই অনন্য বৈশ্বিক সমাধান থাকে যা নিখুঁতভাবে নির্ণয় করা সম্ভব।

৩. ল্যাগ্রাঞ্জিয়ান ডুয়াল ও কেকেটি শর্ত

শর্তযুক্ত সমস্যা সমাধানের জন্য আমরা ল্যাগ্রাঞ্জ গুণক $alpha_i ge 0$ যুক্ত করে ডুয়াল সমীকরণে রূপান্তর করি:

mathcal{L}(w, b, alpha) = frac{1}{2} |w|^2 – sum_{i=1}^n alpha_i left[y_i(w cdot x_i + b) – 1 right]

প্রথম ডেরিভেটিভ শূন্য ধরে পাই $w = sum_{i=1}^n alpha_i y_i x_i$ এবং $sum alpha_i y_i = 0$। এগুলো ল্যাগ্রাঞ্জিয়ানে প্রতিস্থাপন করলে পাওয়া যায় সুবিখ্যাত **ডুয়াল অপ্টিমাইজেশন সমস্যা**:

max_alpha sum_{i=1}^n alpha_i – frac{1}{2} sum_{i,j} alpha_i alpha_j y_i y_j (x_i cdot x_j) quad text{শর্ত:} quad alpha_i ge 0, quad sum alpha_i y_i = 0

লক্ষ করুন, ডুয়াল সমীকরণে ডেটা পয়েন্টগুলো কেবল পরস্পরের **ডট প্রোডাক্ট** ($x_i cdot x_j$) হিসেবে উপস্থিত থাকে—যা পরবর্তীতে কার্নেল ট্রিকের জন্ম দেয়!

৪. স্পার্স সৌন্দর্যের প্রতীক: সাপোর্ট ভেক্টর

কারুশ-কুন-টাকার (KKT) কমপ্লিমেন্টারি স্ল্যাকনেস শর্তটি হলো:

alpha_i left[y_i(w cdot x_i + b) – 1 right] = 0

এর এক গভীর জ্যামিতিক তাৎপর্য রয়েছে:

- যেসব ডেটা পয়েন্ট মার্জিনের বাইরে নিরাপদে অবস্থান করছে (y_i(w cdot x_i + b) > 1), তাদের জন্য $alpha_i = 0$ **হতে বাধ্য**।
- কেবলমাত্র যেসব পয়েন্ট সরাসরি মার্জিনের সীমানায় অবস্থান করছে ($y_i(wcdot x_i + b) = 1$), তাদের জন্যই alpha_i > 0 হতে পারে।

এই সীমানার বিন্দুগুলোকেই বলা হয় **সাপোর্ট ভেক্টর (Support Vectors)**। যেহেতু $w = sum alpha_i y_i x_i$, তাই সম্পূর্ণ সিদ্ধান্ত রেখাটি নির্ভর করে কেবল এই অল্প কয়েকটি সাপোর্ট ভেক্টরের ওপর! বাকি সব ডেটা মুছে ফেললেও হাইপারপ্লেন বিন্দুমাত্র পরিবর্তিত হবে না।

Shakil Mahmud • Research Monograph

৫. সফট মার্জিন ও স্ল্যাক ভেরিয়েবল

বাস্তব ডেটায় কিছুটা নয়েজ থাকতে পারে। এর জন্য স্ল্যাক ভেরিয়েবল xi_ige0 এবং পেনাল্টি প্যারামিটার C যুক্ত করা হয়, যা ডুয়াল সমীকরণে রূপ নেয় একটি বক্স কনস্ট্রেইন্টে: $0lealpha_ileC$। মার্জিনের এই অনন্য গাণিতিক বিশুদ্ধতার কারণেই প্রায় দুই দশক ধরে সাপোর্ট ভেক্টর মেশিন গোটা বিশ্বের মেশিন লার্নিং গবেষণায় অবিসংবাদিত রাজত্ব করেছিল।



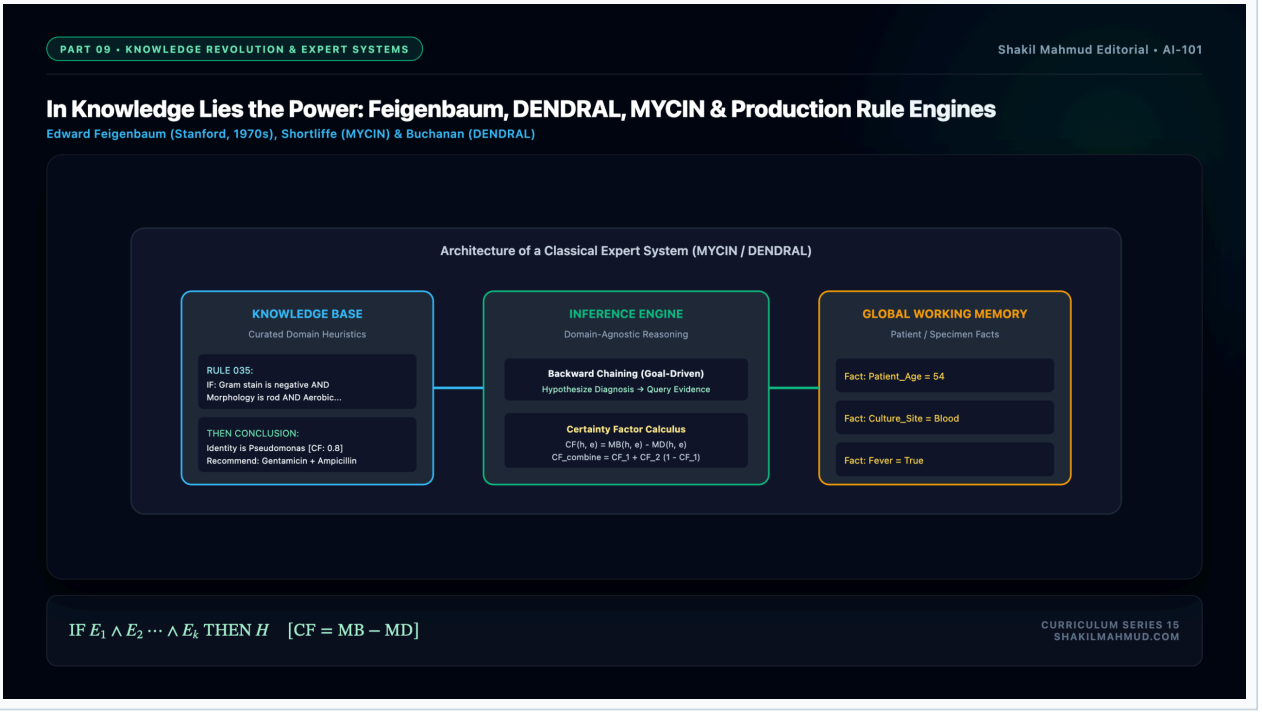
সাইটেশন: মাহমুদ, শাকিল (2026)। "মার্জিনের জ্যামিতি — সাপোর্ট ভেক্টর মেশিন, কেকেটি শর্তাবলি এবং সাপোর্ট ভেক্টর"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ৮।
https://shakilmahmud.com/bn/foundations-of-machine-learning-part-08-geometry-of-margin-support-vector-machines-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • নবম খণ্ড

রিপ্রোডিউসিং কার্নেল হিলবার্ট স্পেস — মার্সারের উপপাদ্য এবং অসীম-মাত্রিক কার্নেল ট্রিক

রিপ্রোডিউসিং কার্নেল হিলবার্ট স্পেস — মার্সারের উপপাদ্য এবং অসীম-মাত্রিক কার্নেল ট্রিক

495 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ৯: *ক্লাসিক্যাল এক্সপার্ট সিস্টেমের গঠন: নলেজ বেস রুলস, ব্যাকওয়ার্ড চেইনিং ইনফারেন্স ইঞ্জিন এবং সার্টেইনটি ফ্যাক্টর ক্যালকুলাস।*

রৈখিক মডেলগুলোর গাণিতিক বিশুদ্ধতা অনন্য হলেও বাস্তব জগতের প্রাকৃতিক ও জৈবিক ঘটনাগুলো চরমভাবে অ-রৈখিক (Non-linear)। চিরাচরিত সমাধান ছিল ইনপুট ডেটাকে বহুপদী বা ত্রিকোণমিতিক সূত্রের মাধ্যমে উচ্চ মাত্রার স্পেসে রূপান্তর করা—কিন্তু তা অচিরেই ডাইমেনশন বিস্ফোরণের (Curse of Dimensionality) প্রাচীরে ধাক্কা খায়। ১৯৯২ সালে বোসার, গায়ন এবং ভ্যাপনিক ১৯০৯ সালের ব্রিটিশ গণিতবিদ জেমস মার্সারের এক বিস্মৃত উপপাদ্য পুনরুজ্জীবিত করে আবিষ্কার করেন **কার্নেল ট্রিক (The Kernel Trick)**: কোনো স্থানাঙ্ক গণনা না করেই অসীম-মাত্রিক হিলবার্ট স্পেসে নিখুঁত রৈখিক সমাধান অর্জনের এক অবিশ্বাস্য কৌশল।

১. উচ্চ মাত্রায় রূপান্তর ও গণনামূলক প্রাচীর

দ্বিমাত্রিক তলে বৃত্তাকারভাবে বিন্যস্ত একটি ডেটাসেট কল্পনা করা যাক, যা কোনো সরলরেখা দিয়ে আলাদা করা অসম্ভব। কিন্তু যদি আমরা একটি ফাংশন $Phi : mathbbR^2 to mathbbR^3$ ব্যবহার করে ডেটাকে ত্রিমাত্রিক তলে তুলে আনি:

$$Phi(x) = left(x_1^2,, sqrt2 x_1 x_2,, x_2^2 right)^T$$

তবে দ্বিমাত্রিক বৃত্তটি ত্রিমাত্রিক স্পেসে একটি সমতল পৃষ্ঠ বা হাইপারপ্লেনে পরিণত হয় এবং ডেটা সম্পূর্ণ রৈখিকভাবে বিভাজ্য হয়ে যায়।

কিন্তু ইনপুট ডেটার ডাইমেনশন যদি হয় $d = 1000$ এবং আমরা ডিগ্রি-৫ বহুপদীতে যেতে চাই, তবে নতুন স্পেসের মাত্রা দাঁড়ায়:

Shakil Mahmud • Research Monograph

$$dim(mathcalH) = binom1000 + 55approx৮textট্রিলিয়নস্থানাঙ্ক!$$

এত বিশাল ভেক্টর মেমরিতে রাখা বা গণনা করা যেকোনো কম্পিউটারের জন্য অসম্ভব।

২. কার্নেল ট্রিক: ডুয়াল ডট প্রোডাক্টের জাদু

এসভিএম-এর ডুয়াল অপ্টিমাইজেশন সমীকরণটি পুনরায় লক্ষ্য করা যাক:

max_alpha sum_{i=1}^n alpha_i – frac{1}{2} sum_{i,j=1}^n alpha_i alpha_j y_i y_j langle Phi(x_i), Phi(x_j) rangle

সবচেয়ে বিস্ময়কর সত্য হলো: **অ্যালগরিদম কখনোই $Phi(x)$-এর স্থানাঙ্কগুলো আলাদাভাবে দাবি করে না**। পুরো শিখন প্রক্রিয়া পরিচালিত হয় কেবল দুটি বিন্দুর ভেতরের ইনার প্রোডাক্ট বা ডট প্রোডাক্ট $langlePhi(x), Phi(z)rangle$-এর মাধ্যমে। যদি আমরা এমন একটি সহজ ফাংশন $K(x, z)$ তৈরি করতে পারি যা মূল ইনপুট স্পেসেই এই ডট প্রোডাক্টের মান হুবহু বের করে দেয়:

$$K(x, z) = langlePhi(x), Phi(z)rangle_{mathcalH}$$

তবে আমাদের কখনোই ৮ ট্রিলিয়ন মাত্রার স্পেসে পা দেওয়ার প্রয়োজন নেই!

যেমন বহুপদী কার্নেল $K(x, z) = (x^T z + c)^p$-এর ক্ষেত্রে মূল d-মাত্রায় সাধারণ একটি ডট প্রোডাক্ট করলেই স্বয়ংক্রিয়ভাবে ট্রিলিয়ন মাত্রার হিলবার্ট স্পেসের ফলাফল পাওয়া যায়।

৩. মার্সারের উপপাদ্য ও আরকেএইচএস (RKHS)

কোন ফাংশনগুলো বৈধ কার্নেল হতে পারে? এর উত্তর দেয় ১৯০৯ সালের **মার্সারের উপপাদ্য (Mercer's Theorem)**:

> ***মার্সারের উপপাদ্য:*** *একটি প্রতিসম অবিচ্ছিন্ন ফাংশন $K(x, z)$ কেবল তখনই একটি বৈধ ইনার প্রোডাক্ট কার্নেল হবে যদি যেকোনো সসীম ডেটা সেটের জন্য গঠিত $ntimesn$ **গ্রাম ম্যাট্রিক্স (Gram Matrix)** $mathbfK_ij = K(x_i, x_j)$ সর্বদা **পজিটিভ সেমি-ডেফিনিট (Positive Semi-Definite)** হয় ($v^T mathbfKvge0$)।*

এর ফলে কার্নেলটিকে তার আইগেনভ্যালু lambda_i > 0 এবং আইগেনফাংশন $psi_i(x)$-এর সাহায্যে অসীম ধারায় প্রকাশ করা সম্ভব হয়:

$$K(x, z) = sum_i=1^infty lambda_i psi_i(x) psi_i(z)$$

৪. গাউসিয়ান আরবিএফ কার্নেল: অসীম মাত্রা

মেশিন লার্নিংয়ের সর্বাধিক জনপ্রিয় কার্নেল হলো **রেডিয়াল বেসিস ফাংশন বা গাউসিয়ান আরবিএফ (RBF) কার্নেল**:

K_{text{RBF}}(x, z) = expleft(-frac{|x – z|^2}{2sigma^2}right)

টেলর সিরিজ দিয়ে এর এক্সপোনেনশিয়াল অংশকে ভাঙলে পাই:

$$exp(2gammax^Tz) = 1 + frac2gamma(x^Tz)1! + frac(2gamma)^2(x^Tz)^22! + dots + frac(2gamma)^k(x^Tz)^kk! + dots$$

যেহেতু এই টেলর সিরিজটি অসীম পর্যন্ত বিস্তৃত, তাই আরবিএফ কার্নেল মূলত ডেটাকে একটি **অসীম-মাত্রিক হিলবার্ট স্পেসে** তুলে নেয়! এর মাধ্যমে যেকোনো জটিল, আঁকাবাঁকা ডেটা ক্লাস্টারকে অত্যন্ত মসৃণভাবে পৃথক করা সম্ভব হয়।

সাইটেশন: মাহমুদ, শাকিল (2026)। "রিপ্রোডিউসিং কার্নেল হিলবার্ট স্পেস — মার্সারের উপপাদ্য এবং অসীম-মাত্রিক কার্নেল ট্রিক"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ৯। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-09-reproducing-kernel-hilbert-space-mercer-theorem-bn/

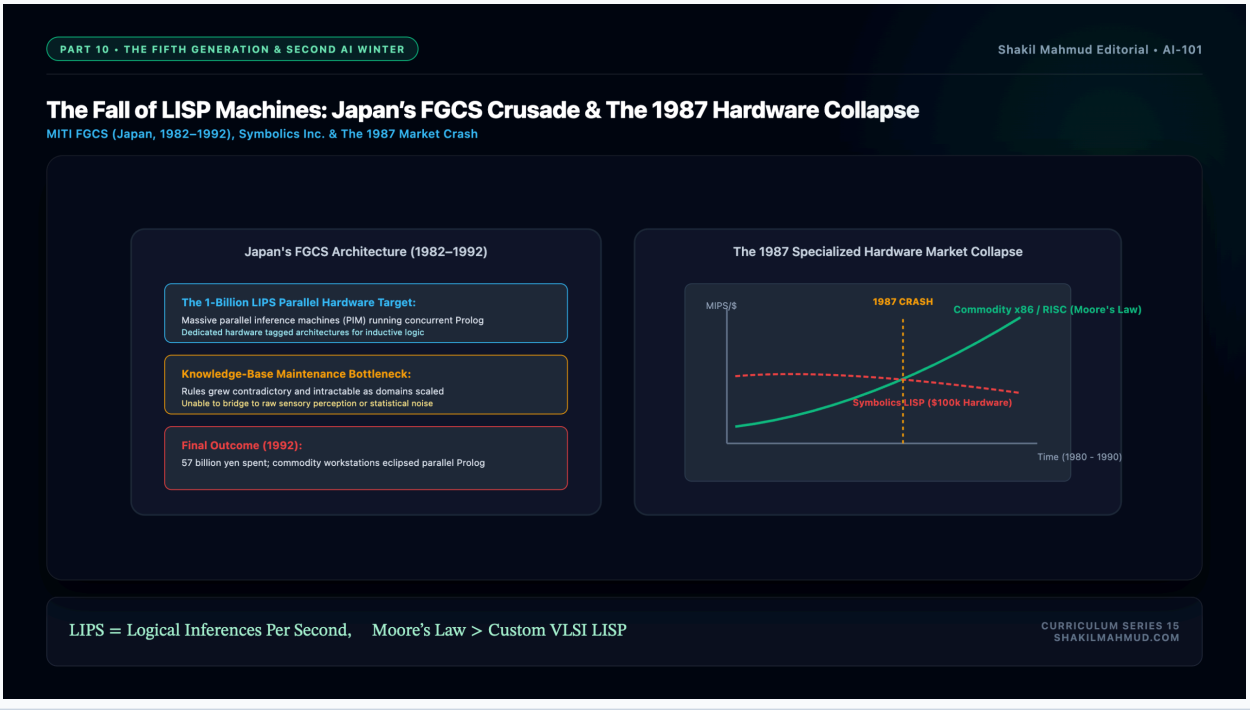
Shakil Mahmud • Research Monograph

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • দশম খণ্ড

স্পার্সিটি ও রেগুলারাইজেশন মেকানিক্স — এল১ ল্যাসো, এল২ রিজ এবং সাব-ডিফারেনশিয়াল জ্যামিতি

স্পার্সিটি ও রেগুলারাইজেশন মেকানিক্স — এল১ ল্যাসো, এল২ রিজ এবং সাব-ডিফারেনশিয়াল জ্যামিতি

483 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১০: *জাপানের পঞ্চম প্রজন্মের প্যারালাল প্রোলগ প্রকল্প এবং ১৯৮৭ সালের লিস্প (LISP) মেশিন ও বিশেষায়িত হার্ডওয়্যার বাজারের আকস্মিক পতন।*

আধুনিক ডেটা বিজ্ঞানে আমরা প্রায়শই এমন এক বাস্তবতার মুখোমুখি হই যেখানে নমুনার তুলনায় ফিচারের সংখ্যা বহুগুণ বেশি ($pggn$)—যেমন ১০০ জন রোগীর ক্ষেত্রে ২০,০০০ জিনের প্রকাশ পরিমাপ। এই চরম পরিস্থিতিতে সাধারণ লিস্ট স্কয়ার্স পদ্ধতিতে অসীম সংখ্যক সমাধান তৈরি হয় যা ট্রেনিং ডেটা নিখুঁত মুখস্থ করলেও টেস্ট ডেটায় পুরোপুরি ব্যর্থ হয়। এই গাণিতিক সংকট নিরসনে ব্যবহৃত হয় **রেগুলারাইজেশন (Regularization)**: মডেলের প্যারামিটারের ওপর পেনাল্টি আরোপ করে অপ্রয়োজনীয় নয়েজ ছাঁটাই করা এবং গুরুত্বপূর্ণ ফিচারগুলোকে স্বয়ংক্রিয়ভাবে শনাক্ত করা।

১. অপূর্ণ-নির্ধারিত সমস্যা ও পেনাল্টির প্রয়োজনীয়তা

লিনিয়ার রিগ্রেশনে $y = Xw + epsilon$, যেখানে $XinmathbbR^{ntimesp}$। যখন p > n, তখন গ্রাম ম্যাট্রিক্স X^TX-এর র‍্যাঙ্ক সর্বোচ্চ n হতে পারে, ফলে এর কোনো ইনভার্স বা বিপরীত ম্যাট্রিক্স থাকে না। প্যারামিটারের ভ্যারিয়েন্স অসীমে বিস্ফোরিত হয়।

রেগুলারাইজেশন এই সমস্যার সমাধান করে অবজেক্টিভ ফাংশনে একটি পেনাল্টি $Omega(w)$ এবং কন্ট্রোল প্যারামিটার lambda > 0 যোগ করে:

$$min_w frac{1}{2n} |y – X w|_2^2 + lambda Omega(w)$$

২. L_2 রিজ রেগুলারাইজেশন (Ridge Regression)

আন্দ্রে তিখোনভ এবং পরবর্তীতে হোয়ার্ল ও কেনার্ড ইউক্লিডীয় দূরত্বের বর্গের পেনাল্টি প্রস্তাব করেন: $Omega(w) = frac12|w|_2^2 = frac12sumw_j^2$।

একে ব্যবকলন করে শূন্য ধরলে এক বিশ্বস্ত সমাধান পাওয়া যায়:

$$hatw_textRidge = left(X^T X + nlambdaI_pright)^-1 X^T y$$

ম্যাট্রিক্সের কর্ণ বরাবর $nlambda$ যোগ করার ফলে সকল আইগেনভ্যালু ধণাত্মক হয়ে যায় এবং ম্যাট্রিক্সটি নিশ্চিতভাবে ইনভার্ট করা সম্ভব হয়। রিজ রিগ্রেশন মূলত একটি অবিচ্ছিন্ন ছাঁকনির মতো কাজ করে যা অপ্রয়োজনীয় ছোট দিকগুলোর মান কমিয়ে দেয়, তবে কোনো প্যারামিটারকে কখনোই সম্পূর্ণ শূন্য করে না।

৩. L_1 ল্যাসো (Lasso) ও স্পার্স ফিচার নির্বাচন

১৯৯৬ সালে রবার্ট তিবশিরানি প্রস্তাব করেন **ল্যাসো (Lasso — Least Absolute Shrinkage and Selection Operator)**, যা প্যারামিটারের পরম মান যোগ করে পেনাল্টি দেয়:

min_w frac{1}{2n} |y – X w|_2^2 + lambda sum_{j=1}^p |w_j|

ল্যাসোর এক বিস্ময়কর গাণিতিক ক্ষমতা রয়েছে: $lambda$-এর মান বৃদ্ধি পেলে এটি অপ্রয়োজনীয় ফিচারগুলোর মান **সম্পূর্ণরূপে শূন্য ($w_j = 0$)** করে দেয়! অর্থাৎ এটি একই সাথে মডেল প্রশিক্ষণ এবং গুরুত্বপূর্ণ ফিচার বাছাইয়ের কাজ সম্পন্ন করে।

৪. স্পার্সিটির জ্যামিতিক উৎস

কেন L_1 শূন্য তৈরি করে কিন্তু L_2 করে না? এর উত্তর লুকিয়ে আছে তাদের সীমাবদ্ধতার জ্যামিতিতে:

- **L_2 রিজের গোলক ($|w|_2^2leC$):** এটি একটি মসৃণ বৃত্তাকার তল। কোয়াড্রেটিক লস ফাংশনের উপবৃত্তগুলো যখন এই বৃত্তকে স্পর্শ করে, তখন তারা কোনো সাধারণ বক্ররেখার বিন্দুতে স্পর্শ করে, যেখানে উভয় প্যারামিটারের মান অশূন্য (w_1ne0, w_2ne0) থাকে।
- **L_1 ল্যাসোর ডায়মন্ড ($|w|_1leC$):** দ্বিমাত্রিক তলে এটি একটি হীরক বা চতুর্ভুজাকার বহুভুজ, যার ধারালো শীর্ষবিন্দুগুলো সরাসরি অক্ষের ওপর অবস্থান করে। কেন্দ্র থেকে প্রসারিত হতে থাকা লস উপবৃত্তগুলো সবচেয়ে আগে এই ধারালো শীর্ষবিন্দুতে ধাক্কা খায়। আর অক্ষের শীর্ষবিন্দুতে ধাক্কা খাওয়ার অর্থই হলো—অন্য স্থানাঙ্কের মান সম্পূর্ণরূপে শূন্য ($w_2 = 0$)!

৫. প্রক্সিমাল গ্রেডিয়েন্ট ও সফট-থ্রেশহোল্ডিং

যেহেতু শূন্য বিন্দুতে $|w|$-এর কোনো সাধারণ ডেরিভেটিভ নেই, তাই ল্যাসো অপ্টিমাইজেশনে ব্যবহৃত হয় **সফট-থ্রেশহোল্ডিং অপারেটর (Soft-Thresholding)**:

hat{w}_j = text{sign}(w_j^{text{OLS}}) max(0, |w_j^{text{OLS}}| – lambda)

যেসব প্যারামিটারের মান $lambda$-এর চেয়ে ছোট, তারা অবিলম্বে শূন্য হয়ে যায়; আর বড় প্যারামিটারগুলো $lambda$ পরিমাণ সংকুচিত হয়। উচ্চ-মাত্রার ডেটা ফিল্টারিংয়ে L_1 রেগুলারাইজেশন আজ বিশ্বজুড়ে এক অপরিহার্য গাণিতিক মানদণ্ড।

Shakil Mahmud • Research Monograph

❦

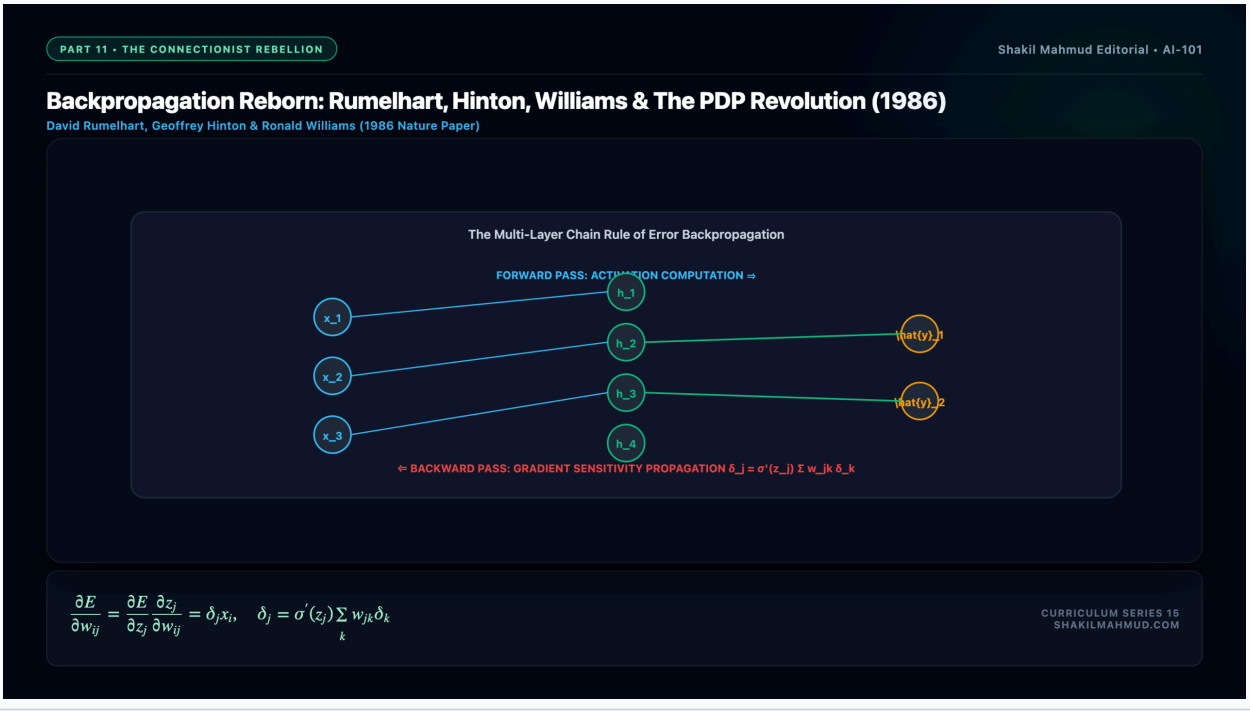
সাইটেশন: মাহমুদ, শাকিল (2026)। "স্পার্সিটি ও রেগুলারাইজেশন মেকানিক্স — এল১ ল্যাসো, এল২ রিজ এবং সাব-ডিফারেনশিয়াল জ্যামিতি"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১০। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-10-sparsity-regularization-mechanics-lasso-ridge-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • একাদশ খণ্ড

ডিসিশন ট্রি ইন্ডাকশন ও রিকার্সিভ পার্টিশনিং — ইনফরমেশন গেইন, জিনি ইমপিউরিটি এবং প্রুনিং

ডিসিশন ট্রি ইন্ডাকশন ও রিকার্সিভ পার্টিশনিং — ইনফরমেশন গেইন, জিনি ইমপিউরিটি এবং প্রুনিং

432 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১১: *মাল্টি-লেয়ার ব্যাকপ্রোপাগেশন অ্যালগরিদম: ফরওয়ার্ড অ্যাক্টিভেশন পাস এবং চেইন রুলের সাহায্যে ব্যাকওয়ার্ড এরর সংবেদনশীলতা বিস্তার।*

যেখানে সাপোর্ট ভেক্টর মেশিন বা রিগ্রেশন মডেল অবিচ্ছিন্ন জ্যামিতিক তল দিয়ে কাজ করে, সেখানে ডিসিশন ট্রি (Decision Tree) পরিচালিত হয় অনুক্রমিক যুক্তির মাধ্যমে। ১৯৮৪ সালে লিও ব্রেইম্যান, জেরোম ফ্রিডম্যান, রিচার্ড ওলশেন এবং চার্লস স্টোন তাঁদের ক্লাসিক মনোগ্রাফ *“Classification and Regression Trees (CART)”*-এ এমন এক কাঠামো তুলে ধরেন যা উচ্চ-মাত্রিক স্পেসকে অক্ষ-বরাবর আয়তক্ষেত্রাকার অঞ্চলে বিভক্ত করে। এর ফলে তৈরি হয় সম্পূর্ণ ব্যাখ্যাকরণযোগ্য এক সিদ্ধান্ত কাঠামো যা স্কেলিং পরিবর্তনের মুখেও চিরস্থির থাকে।

১. অর্থোগোনাল স্পেস পার্টিশনিংয়ের জ্যামিতি

ধরা যাক পুরো ইনপুট স্পেস $mathcal X$-কে M সংখ্যক অসদৃশ অঞ্চলে $R_1, dots, R_M$ বিভক্ত করা হলো। একটি ডিসিশন ট্রি পূর্বাভাস দেয় একটি পিস-ওয়াইজ ধ্রুবক ফাংশন হিসেবে:

$$f(x) = sum_{m=1}^{M} c_m mathbb{I}(x in R_m)$$

যেখানে ক্লাসিফিকেশনের ক্ষেত্রে c_m হলো ওই অঞ্চলের সর্বাধিক সংখ্যাগরিষ্ঠ ক্লাসের লেবেল।

গাণিতিকভাবে d মাত্রায় M সংখ্যক সর্বোত্তম অঞ্চলের সীমানা নিখুঁতভাবে বের করা একটি এনপি-কমপ্লিট (NP-complete) সমস্যা। তাই ট্রি তৈরি হয় গ্রিডি বা লোভী রিকার্সিভ বাইনারি স্প্লিটিংয়ের মাধ্যমে।

Shakil Mahmud • Research Monograph

২. নোড অপবিত্রতা পরিমাপ: এনট্রপি বনাম জিনি

প্রতিটি নোডে ডেটা কতটা বিশৃঙ্খল বা মিশ্রিত তা পরিমাপ করতে **ইমপিউরিটি ফাংশন** $I(m)$ ব্যবহৃত হয়:

1. **শ্যানন এনট্রপি (Shannon Entropy):** তথ্য তত্ত্বের আলোকে ডেটার অনিশ্চয়তা পরিমাপ করে:

$$I_E(m) = -sum_{k=1}^{K} p_{mk} log_2 p_{mk}$$

   এনট্রপির হ্রাসকে বলা হয় **ইনফরমেশন গেইন (Information Gain)**।

2. **জিনি ইমপিউরিটি (Gini Impurity — CART):** একটি নোড থেকে দৈবভাবে একটি উপাদান তুলে নিয়ে তাকে নোডের ডিস্ট্রিবিউশন অনুযায়ী লেবেল দিলে তা ভুল হওয়ার সম্ভাব্যতা:

I_G(m) = 1 – sum_{k=1}^K p_{mk}^2

   যদি নোডে কেবল এক ক্লাসের ডেটা থাকে (বিশুদ্ধ নোড), তবে জিনির মান শূন্য হয়।

3. **মিন স্কয়ার্ড এরর (MSE):** রিগ্রেশন ট্রির ক্ষেত্রে মানের ভ্যারিয়েন্স পরিমাপ করে।

৩. গ্রিডি স্প্লিটিং নীতি

প্রতিটি ধাপে অ্যালগরিদম সকল ফিচারের সম্ভাব্য সকল কাট-পয়েন্ট s পরীক্ষা করে নোডটিকে ডানে ও বামে বিভক্ত করে:

R_L(j, s) = {x mid x_j le s}, quad R_R(j, s) = {x mid x_j > s}

যে স্প্লিটটি গড় ইমপিউরিটি সর্বাধিক হ্রাস করতে পারে, সেটিকেই চূড়ান্ত হিসেবে গ্রহণ করা হয়:

Delta I = I(m) – left(frac{|D_L|}{|D_m|} I(D_L) + frac{|D_R|}{|D_m|} I(D_R) right)

৪. ওভারফিটিং ও কস্ট-কমপ্লেক্সিটি প্রুনিং (Pruning)

কোনো বাধা না দিলে একটি ট্রি ততক্ষণ পর্যন্ত ডালপালা ছড়াতে থাকে যতক্ষণ না প্রতিটি পাতায় মাত্র একটি করে ডেটা অবশিষ্ট থাকে ($I(m) = 0$)। এর ফলে ট্রেনিং এরর শূন্য হলেও ভ্যারিয়েন্স চরমভাবে বৃদ্ধি পায় এবং মডেলটি ব্যর্থ হয়।

কার্ট (CART) এই সমস্যা সমাধান করে এক অনন্য কৌশলে: প্রথমে কোনো বাধা ছাড়াই একটি সুবিশাল ট্রি T_0 বড় হতে দেওয়া হয়, এবং পরে পেছন থেকে অপ্রয়োজনীয় ডালপালা ছাঁটাই বা **প্রুনিং (Pruning)** করা হয়। এর জন্য **কস্ট-কমপ্লেক্সিটি শর্ত** ব্যবহার করা হয়:

$$mathcal{C}_alpha(T) = sum_{m in textleaves(T)} |D_m| I(m) + alpha |textleaves(T)|$$

এখানে $alpha$ হলো ট্রির আকারের ওপর একটি পেনাল্টি। ক্রস-ভ্যালিডেশনের মাধ্যমে সঠিক $alpha^*$ নির্বাচন করে ট্রি ছাঁটাই করার মাধ্যমে বায়াস ও ভ্যারিয়েন্সের নিখুঁত ভারসাম্য নিশ্চিত করা হয়।

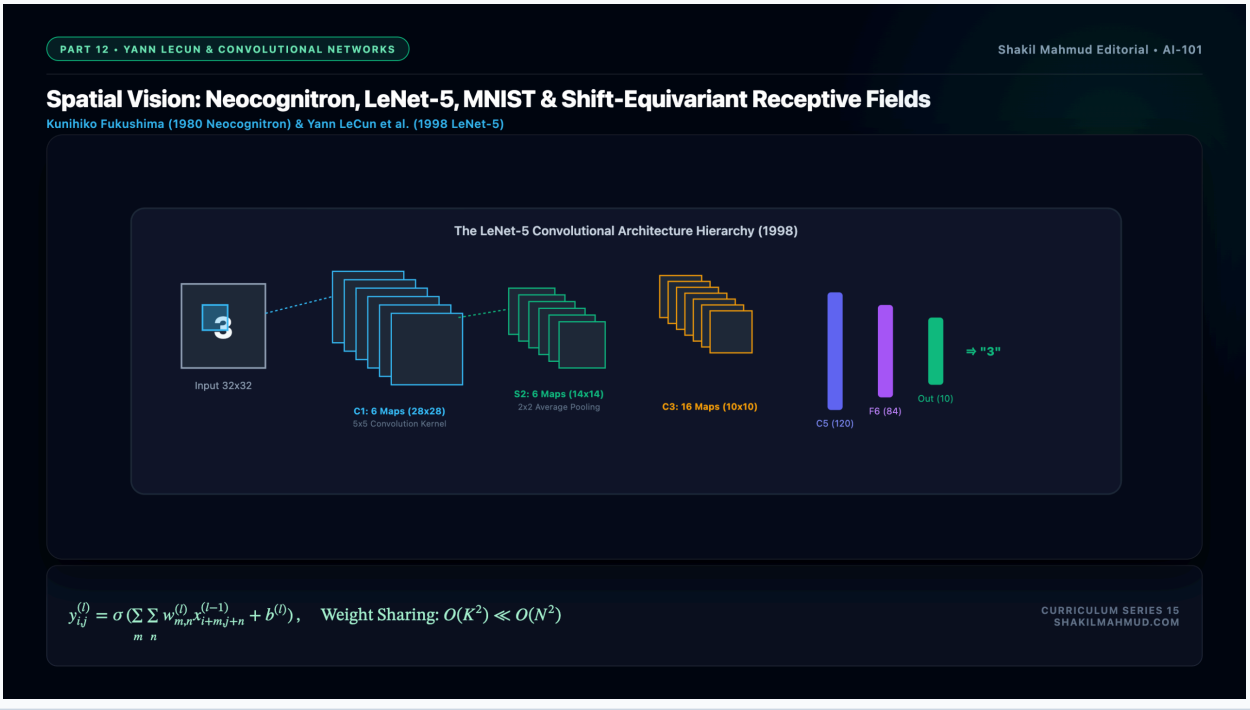
সাইটেশন: মাহমুদ, শাকিল (2026)। "ডিসিশন ট্রি ইন্ডাকশন ও রিকার্সিভ পার্টিশনিং — ইনফরমেশন গেইন, জিনি ইমপিউরিটি এবং প্রুনিং"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১১। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-11-tree-induction-recursive-partitioning-cart-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • দ্বাদশ খণ্ড

এনসেম্বল লার্নিং ও ভ্যারিয়েন্স রিডাকশন — বুটস্ট্র্যাপ অ্যাগ্রিগেশন, ওওবি বাউন্ড এবং র‍্যান্ডম ফরেস্ট

এনসেম্বল লার্নিং ও ভ্যারিয়েন্স রিডাকশন — বুটস্ট্র্যাপ অ্যাগ্রিগেশন, ওওবি বাউন্ড এবং র‍্যান্ডম ফরেস্ট

511 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১২: *ইয়ান লেকানের লেনাট-৫ (LeNet-5) কনভোলিউশনাল নেটওয়ার্ক: লোকাল রিসেপ্টিভ ফিল্ড, ওজন শেয়ারিং এবং সাবস্যাম্পলিং স্তরবিন্যাস।*

ডিসিশন ট্রির প্রকাশক্ষমতা এবং শেখার গতি অত্যন্ত দ্রুত হলেও এর একটি মারাত্মক দুর্বলতা ছিল: চরম অস্থিরতা। ট্রেনিং সেটের সামান্য একটি ডেটা পয়েন্ট পরিবর্তন করলে মূল স্প্লিটের সিদ্ধান্ত বদলে যেত, যার ফলে নিচের সম্পূর্ণ গাছটির কাঠামো উল্টেপাল্টে যেত। ১৯৯৬ এবং ২০০১ সালে মার্কিন গণিতবিদ লিও ব্রেইম্যান পরিসংখ্যানিক গড়ের শক্তি ব্যবহার করে এই দুর্বলতাকেই শ্রেষ্ঠ অস্ত্রে রূপান্তরিত করেন। তিনি আবিষ্কার করেন **বুটস্ট্র্যাপ অ্যাগ্রিগেশন (Bagging)** এবং **র‍্যান্ডম ফরেস্ট (Random Forest)**—যা আজও ট্যাবুলার ডেটা বিশ্লেষণে পৃথিবীর সবচেয়ে শক্তিশালী অ্যালগরিদম।

১. এনসেম্বল গড়ের ভ্যারিয়েন্স বিশ্লেষণ

ধরা যাক আমাদের কাছে B সংখ্যক আলাদা প্রেডিক্টর রয়েছে, যাদের প্রত্যেকের নিজস্ব ভ্যারিয়েন্স $sigma^2$ এবং তাদের মধ্যকার পারস্পরিক কোরিলেশন rho। যদি আমরা এদের গড় নিয়ে একটি সম্মিলিত এনসেম্বল $barf = frac1Bsumhatf_b$ তৈরি করি, তবে তার মোট ভ্যারিয়েন্স দাঁড়ায়:

$$text{Var}(bar{f}) = rho sigma^2 + frac{1 – rho}{B} sigma^2$$

এনসেম্বলের আকার যখন অসীমের দিকে যায় ($Btoinfty$):

Shakil Mahmud • Research Monograph

$$lim_{Btoinfty} textVar(barf) = rhosigma^2$$

এই সমীকরণটি এনসেম্বল শিখনের মূল সত্যটি প্রকাশ করে দেয়:

- ডানদিকের অংশ frac{1 – rho}{B} sigma^2 গাছ বা মডেলের সংখ্যা বাড়িয়ে অনায়াসেই শূন্যে নামিয়ে আনা যায়।
- কিন্তু ভ্যারিয়েন্স কখনোই **অ্যাসিম্পটোটিক ফ্লোর** $rhosigma^2$-এর নিচে নামতে পারে না! অর্থাৎ, গাছগুলো যদি নিজেদের মধ্যে একই রকম হয় ($rho = 1$), তবে হাজারটা গাছ যোগ করলেও ভ্যারিয়েন্স বিন্দুমাত্র কমবে না।

সুতরাং সফল এনসেম্বল তৈরির মূল রহস্য হলো—গাছগুলোর পারস্পরিক কোরিলেশন rho-কে শূন্যের কাছাকাছি নামিয়ে আনা।

২. বুটস্ট্র্যাপ অ্যাগ্রিগেশন বা ব্যাগিং (Bagging)

একটি একক ডেটাসেট S থেকে বহু স্বাধীন গাছ তৈরি করতে ব্রেইম্যান ব্যবহার করেন ব্র্যাডলি এফরনের **বুটস্ট্র্যাপ (Bootstrap)** পদ্ধতি: প্রতিস্থাপনের সুযোগ রেখে (with replacement) মূল ডেটাসেট থেকে নতুন n আকারের নমুনা তৈরি করা।

গাণিতিক সম্ভাব্যতা অনুযায়ী, বুটস্ট্র্যাপ স্যাম্পলিংয়ে যেকোনো ডেটা পয়েন্টের বাদ পড়ার সম্ভাবনা হলো:

lim_{n to infty} left(1 – frac{1}{n} right)^n = frac{1}{e} approx ৩৬.৮%

অর্থাৎ প্রতিটি বুটস্ট্র্যাপ সেটে প্রায় ৬৩.২% মৌলিক ডেটা থাকে এবং বাকি ৩৬.৮% ডেটা ব্যবহৃত হয় না। এই অব্যবহৃত অংশকে বলা হয় **আউট-অব-ব্যাগ ডেটা (OOB — Out-of-Bag)**। এটি একটি অভ্যন্তরীণ ভ্যালিডেশন সেট হিসেবে কাজ করে, ফলে আলাদা টেস্ট ডেটা ছাড়াই মডেলের নির্ভুলতা নিরপেক্ষভাবে পরিমাপ করা সম্ভব হয়।

৩. র‍্যান্ডম ফরেস্ট: কোরিলেশন ভাঙার মাস্টারস্ট্রোক

ব্যাগিং ভ্যারিয়েন্স কমালেও গাছগুলোর মধ্যকার কোরিলেশন তখনও বেশ বেশি থাকত ($rhoapprox0.6$)। কারণ ডেটাতে কোনো একটি শক্তিশালী ফিচার থাকলে প্রতিটি গাছ শুরুতেই সেই ফিচার দিয়ে ভাগ করত।

২০০১ সালে ব্রেইম্যান নিয়ে আসেন **র‍্যান্ডম ফরেস্ট (Random Forest)**। তিনি প্রতিটি নোডে ফিচার বাছাইয়ের ক্ষেত্রেও দৈবতা যুক্ত করেন:

1. প্রতিটি স্প্লিটে সকল p সংখ্যক ফিচার না দেখে কেবল এলোমেলোভাবে বাছাই করা $mapproxsqrtp$ সংখ্যক ফিচার বিবেচনা করা হয়।
2. কেবল সেই ক্ষুদ্র উপসেটের মধ্য থেকেই সেরা স্প্লিটটি নির্বাচন করা হয়।

এর ফলে কোনো একক ফিচার পুরো ফরেস্টকে নিয়ন্ত্রণ করতে পারে না। প্রতিটি গাছ বাধ্য হয় ভিন্ন ভিন্ন ফিচারের সম্পর্ক আবিষ্কার করতে। এর ফলে পারস্পরিক কোরিলেশন rho এক ধাক্কায় ০.৬ থেকে ০.১৫-তে নেমে আসে, এবং ফরেস্টের সামগ্রিক ভ্যারিয়েন্স নাটকীয়ভাবে হ্রাস পায়।

৪. ফরেস্ট হরাইজনের বিশেষ বৈশিষ্ট্য

র‍্যান্ডম ফরেস্টের দুটি অনন্য বৈশিষ্ট্য রয়েছে:

- **গাছ বাড়ালে ওভারফিটিং হয় না:** বৃহদাকার সংখ্যার সূত্রের কারণে $Btoinfty$ হলে ফরেস্টের ত্রুটি একটি সুনির্দিষ্ট সীমায় স্থির হয়ে যায়। হাজার হাজার অতিরিক্ত গাছ যোগ করলেও মডেল কখনো অতিরিক্ত জটিল হয়ে ওভারফিট করে না।
- **বহুমাত্রিক অভিযোজন:** এটি স্বয়ংক্রিয়ভাবে গুরুত্বপূর্ণ ফিচার শনাক্ত করতে পারে এবং স্কেলিংয়ের তারতম্য অগ্রাহ্য করে নিখুঁত পূর্বাভাস দিতে পারে।

সাইটেশন: মাহমুদ, শাকিল (2026)। "এনসেম্বল লার্নিং ও ভ্যারিয়েন্স রিডাকশন — বুটস্ট্র্যাপ অ্যাগ্রিগেশন, ওওবি বাউন্ড এবং র‍্যান্ডম ফরেস্ট"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১২। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-12-variance-reduction-ensembles-random-forests-bn/

Shakil Mahmud • Research Monograph

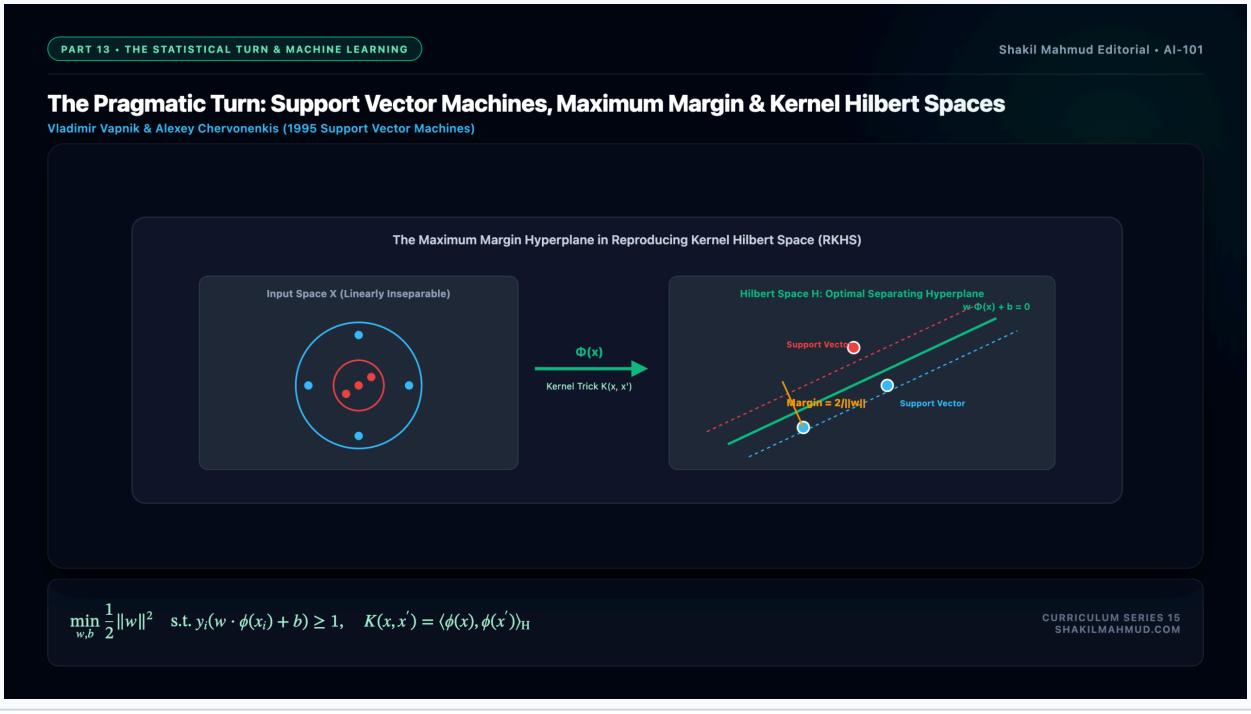
শাকিল মাহমুদ মনোগ্রাফ সিরিজ • ত্রয়োদশ খণ্ড

ফাংশনাল স্পেসে বুস্টিং — অ্যাডাবুস্ট, মার্জিন বাউন্ড এবং এক্সজিবিউস্ট

ফাংশনাল স্পেসে বুস্টিং — অ্যাডাবুস্ট, মার্জিন বাউন্ড এবং এক্সজিবিউস্ট

468 শব্দ • November 4, 2026 • ওয়েব সংস্করণ

❖



চিত্র ১৩: *রিপ্রোডিউসিং কার্নেল হিলবার্ট স্পেসে সাপোর্ট ভেক্টর মেশিন: ম্যাক্সিমাম মার্জিন হাইপারপ্লেন এবং নন-লিনিয়ার কার্নেল ট্রিক।*

১৯৮৮ সালে মাইকেল কার্নস এবং লেসলি ভ্যালিয়ান্ট একটি মৌলিক দার্শনিক প্রশ্ন তুলেছিলেন: এমন কিছু দুর্বল অনুমানকারী (Weak Learners)—যারা মুদ্রা টসের চেয়ে মাত্র সামান্য একটু ভালো কাজ করে (text{Error} < 1/2 – gamma)—তাদের কি সম্মিলিতভাবে যুক্ত করে যেকোনো কাঙ্ক্ষিত উচ্চ-মানের শক্তিশালী মডেলে রূপান্তর করা সম্ভব? ১৯৯০ সালে রবার্ট শ্যাপায়ার প্রমাণ করেন যে হ্যাঁ, এটি তাত্ত্বিকভাবে সত্য। জন্ম হয় **বুস্টিং (Boosting)** ধারণার। পরবর্তীতে জেরোম ফ্রিডম্যান বুস্টিংকে ফাংশনাল স্পেসে গ্রেডিয়েন্ট ডিসেন্টের রূপ দেন, যা আধুনিক ডেটা সায়েন্সের অবিসংবাদিত ভিত্তি তৈরি করে।

১. অ্যাডাবুস্ট (AdaBoost) অ্যালগরিদম

১৯৯৫ সালে ইয়োভ ফ্রয়েন্ড এবং রবার্ট শ্যাপায়ার তৈরি করেন **অ্যাডাবুস্ট (Adaptive Boosting)**। এর মূল কৌশল হলো ভুল করা ডেটা পয়েন্টের ওপর ক্রমান্বয়ে গুরুত্ব বৃদ্ধি করা:

1. শুরুতে সকল ডেটা পয়েন্টের ওজন সমান থাকে ($D_1(i) = 1/n$)।
2. প্রতি ধাপে একটি দুর্বল ডিসিশন ট্রি h_t তৈরি করা হয় এবং তার ওজনযুক্ত ভুল $epsilon_t$ গণনা করা হয়।
3. মডেলটির নির্ভুলতার ওপর ভিত্তি করে তার ভোটিং ক্ষমতা $alpha_t$ নির্ধারিত হয়:

$$alpha_t = frac{1}{2} ln left(frac{1 – epsilon_t}{epsilon_t} right)$$

4. যেসব ডেটায় মডেলটি ভুল করেছে, পরবর্তী ধাপের জন্য তাদের গুরুত্ব সূচকীয় হারে বাড়িয়ে দেওয়া হয়:

$$D_{t+1}(i) propto D_t(i) exp(-alpha_t y_i h_t(x_i))$$

চূড়ান্ত সিদ্ধান্ত গৃহীত হয় সকল দুর্বল মডেলের ওজনযুক্ত ভোটের মাধ্যমে। ফ্রয়েন্ড এবং শ্যাপায়ার প্রমাণ করেন যে ট্রেনিং এরর প্রতি ধাপে সূচকীয় গতিতে শূন্যের দিকে ধাবিত হয়।

২. ওভারফিটিংয়ের বিরুদ্ধে মার্জিনের প্রাচীর

অ্যাডাবুস্টের প্রথমদিকের পরীক্ষায় গবেষকরা এক অবিশ্বাস্য রহস্যের মুখোমুখি হন: ট্রেনিং এরর শূন্য হয়ে যাওয়ার পরেও যদি মডেলকে শত শত অতিরিক্ত রাউন্ড প্রশিক্ষণ দেওয়া হতো, তবুও মডেলটি ওভারফিট করত না; বরং টেস্ট এরর আরও কমতে থাকত! ১৯৯৮ সালে শ্যাপায়ার ও বার্টলেট **মার্জিন তত্ত্ব (Margin Theory)** দিয়ে এই ধাঁধার সমাধান করেন।

তাঁরা দেখান যে ট্রেনিং এরর শূন্য হওয়াটাই শেষ কথা নয়। পরবর্তী রাউন্ডগুলোতে অ্যাডাবুস্ট সবচেয়ে কঠিন ডেটা পয়েন্টগুলোকে সিদ্ধান্ত রেখা থেকে অনেক দূরে ঠেলে দেয়, যার ফলে ডেটাগুলোর **জ্যামিতিক মার্জিন বহুগুণ বৃদ্ধি পায়**। এই বর্ধিত মার্জিন মডেলটিকে নয়েজের বিরুদ্ধে এক অভেদ্য সুরক্ষা দেয়।

৩. ফাংশনাল স্পেসে গ্রেডিয়েন্ট বুস্টিং

২০০১ সালে জেরোম ফ্রিডম্যান বুস্টিংকে এক নতুন দিগন্তে নিয়ে যান: **গ্রেডিয়েন্ট বুস্টিং (Gradient Boosting)**। প্যারামিটারের সাপেক্ষে গ্রেডিয়েন্ট না নিয়ে তিনি সরাসরি মডেলের পূর্বাভাসের সাপেক্ষে লস ফাংশনের নেগেটিভ গ্রেডিয়েন্ট গণনা করেন, যাকে বলা হয় **সিউডো-রেসিডুয়াল (Pseudo-Residuals)**:

$$r_{im} = -left[frac partial L(y_i, F(x_i)) partial F(x_i) right]_{F=F_{m-1}}$$

প্রতি ধাপে নতুন একটি ট্রি তৈরি করা হয় যা পূর্ববর্তী মডেলের এই অবশিষ্ট ভুলের (Residuals) পূর্বাভাস দিতে চেষ্টা করে। এর ফলে যেকোনো ডিফারেনশিয়েবল লস ফাংশনে (MSE, Logistic, Huber) অনায়াসে বুস্টিং প্রয়োগ করা সম্ভব হয়।

৪. এক্সজিবিউস্ট (XGBoost): দ্বিতীয় ক্রমের জ্যামিতি

২০১৬ সালে তিয়ানকি চেন উদ্ভাবন করেন **এক্সজিবিউস্ট (XGBoost)**। তিনি কেবল প্রথম গ্রেডিয়েন্টেই থামেননি, বরং টেলর সিরিজের সাহায্যে হেসিয়ান বা দ্বিতীয় বক্রতা ($h_i = partial^2 L$) যুক্ত করেন:

$$mathcal L approx sum left[g_i f_t(x_i) + frac 12 h_i f_t^2(x_i) right] + Omega(f_t)$$

এর ফলে ট্রির প্রতিটি পাতার সর্বোত্তম ওজন এক নিখুঁত অ্যালজেব্রিক সূত্রে নির্ণয় করা সম্ভব হয়:

$$w_j^* = -frac sum g_i sum h_i + lambda$$

প্রথম গ্রেডিয়েন্টের গতি এবং দ্বিতীয় হেসিয়ানের বক্রতার এই সমন্বয় এক্সজিবিউস্টকে পরিণত করেছে আজকের মেশিন লার্নিং প্রতিযোগিতার সর্বাধিক বিজয়ী ইঞ্জিনে।

Shakil Mahmud • Research Monograph

❦

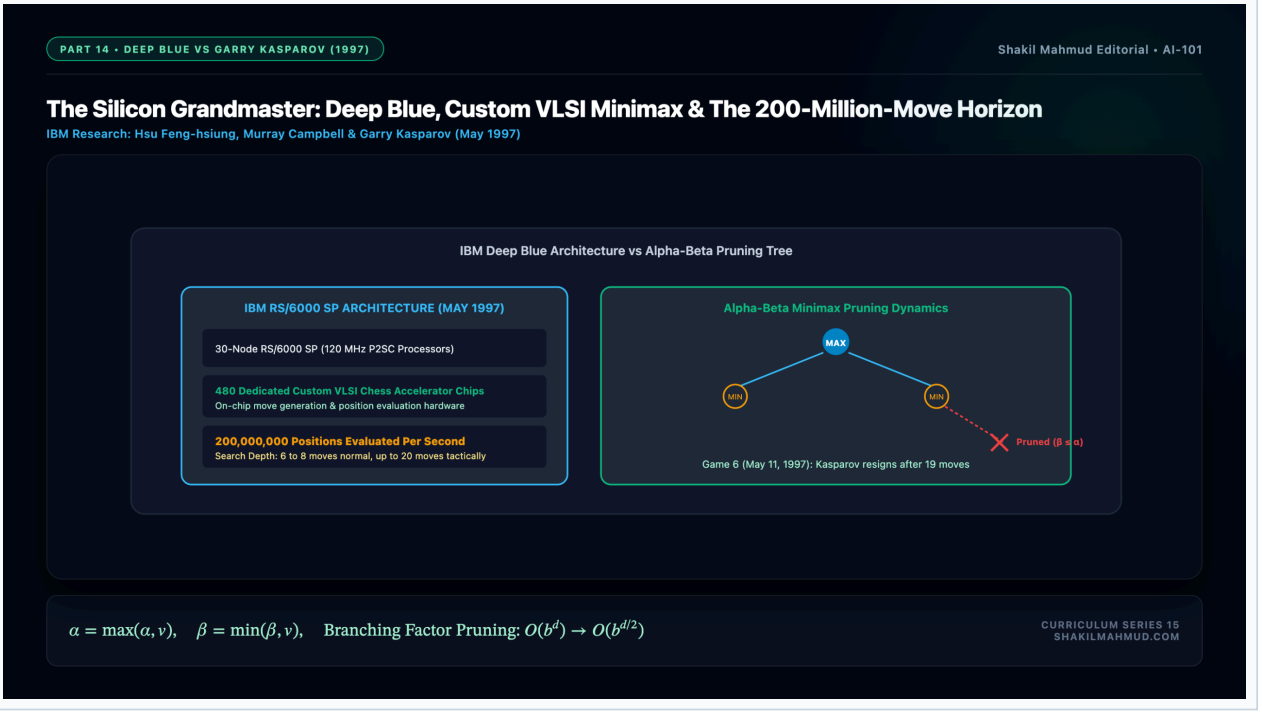
সাইটেশন: মাহমুদ, শাকিল (2026)। "ফাংশনাল স্পেসে বুস্টিং — অ্যাডাবুস্ট, মার্জিন বাউন্ড এবং এক্সজিবিউস্ট"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১৩।
https://shakilmahmud.com/bn/foundations-of-machine-learning-part-13-boosting-function-space-adaboost-xgboost-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • চতুর্দশ খণ্ড

সম্ভাবনাময় গ্রাফিক্যাল মডেল — বায়েসিয়ান নেটওয়ার্ক, এমআরএফ এবং ডি-সেপারেশন

সম্ভাবনাময় গ্রাফিক্যাল মডেল — বায়েসিয়ান নেটওয়ার্ক, এমআরএফ এবং ডি-সেপারেশন

422 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১৪: *আইবিএম ডিপ ব্লুর ৪৮০টি কাস্টম ভিএলএসআই দাবা চিপ এবং কাসপারভের বিরুদ্ধে প্রতি সেকেন্ডে ২০ কোটি অবস্থানের আলফা-বিটা মিনিম্যাক্স সার্চ।*

উচ্চ-মাত্রার স্পেসে সম্ভাব্যতা পরিমাপ করতে গেলে এক মারাত্মক গাণিতিক সংকটের সৃষ্টি হয়: N সংখ্যক বাইনারি ভেরিয়েবলের যৌথ বণ্টন নির্ণয় করতে 2^N – 1 সংখ্যক স্বাধীন প্যারামিটারের প্রয়োজন হয়, যা মাত্র ৫০টি ভেরিয়েবলেই ট্রিলিয়ন ছাড়িয়ে যায়। ১৯৮০-এর দশকের শেষভাগে জুদিয়া পার্ল এবং স্টিফেন লরিটজেন গ্রাফ থিওরি এবং সম্ভাব্যতা তত্ত্বকে একত্রিত করে আবিষ্কার করেন **সম্ভাব্যতামূলক গ্রাফিক্যাল মডেল (Probabilistic Graphical Models — PGM)**। গ্রাফের সংযোগরেখার মাধ্যমে শর্তাধীন স্বাধীনতার সম্পর্ক স্থাপন করে এই মডেলগুলো অসম্ভব জয়েন্ট ডিস্ট্রিবিউশনকে এক সংহত, গুণনীয়ক আকারে প্রকাশ করতে সক্ষম হয়।

১. ডিরেক্টেড গ্রাফ: বায়েসিয়ান নেটওয়ার্ক

একটি **বায়েসিয়ান নেটওয়ার্ক (Bayesian Network)** হলো একটি ডিরেক্টেড অ্যাসাইক্লিক গ্রাফ (DAG), যেখানে প্রতিটি নোড একটি দৈব চলক নির্দেশ করে এবং নির্দেশক তীরগুলো নির্ভরতা নির্দেশ করে। এর সামগ্রিক বণ্টন পিতা-নোডের ওপর ভিত্তি করে গুণফল আকারে ভেঙে যায়:

$$P(X_1, dots, X_N) = prod_{i=1}^{N} P(X_i midtext Parents(X_i))$$

যদি প্রতিটি নোডের সর্বোচ্চ k সংখ্যক পিতা থাকে, তবে প্যারামিটার সংখ্যা সূচকীয় $O(2^N)$ থেকে এক ধাক্কায় পলিনমিয়াল $O(Ncdot 2^k)$-তে নেমে আসে।

Shakil Mahmud • Research Monograph

২. জুদিয়া পার্লের ডি-সেপারেশন (D-Separation)

একটি গ্রাফ থেকে কীভাবে বুঝব যে ভেরিয়েবল A এবং B শর্তাধীনভাবে স্বাধীন কি না? ১৯৮৮ সালে জুদিয়া পার্ল প্রণয়ন করেন **ডি-সেপারেশন (d-separation)** নিয়মাবলি:

1. **কজাল চেইন ($XtoYtoZ$):** Y অজানা থাকলে X ও Z-এর মধ্যে তথ্য প্রবাহিত হয়। কিন্তু Y-কে পর্যবেক্ষণ করলে পথটি বন্ধ হয়ে যায় ($XperpZmidY$)।
2. **সাধারণ কারণ বা ফর্ক ($XleftarrowYtoZ$):** Y হলো একটি যৌথ কারণ। Y-এর মান জানা থাকলে X ও Z একে অপরের থেকে সম্পূর্ণ স্বাধীন হয়ে যায়।
3. **কলাইডার বা যৌথ ফল ($XtoYleftarrowZ$):** এটি সবচেয়ে চমকপ্রদ। স্বাভাবিক অবস্থায় X ও Z সম্পূর্ণ স্বাধীন ($XperpZ$)। কিন্তু যদি তাদের যৌথ ফলাফল Y-কে পর্যবেক্ষণ করা হয়, তবে উল্টো X ও Z-এর মধ্যে সম্পর্ক তৈরি হয় (বার্কসনের প্যারাডক্স বা Explaining Away)!

যদি দুটি নোডের মধ্যকার সকল পথ কোনো না কোনোভাবে ব্লকড বা বিচ্ছিন্ন থাকে, তবে তারা গাণিতিকভাবে সম্পূর্ণ স্বাধীন।

৩. আনডিরেক্টেড গ্রাফ: মার্কভ র‍্যান্ডম ফিল্ডস (MRF)

কম্পিউটার ভিশনের পিক্সেল বা পদার্থবিজ্ঞানের কণার মতো যেখানে কোনো একমুখী কার্যকারণ নেই, সেখানে ব্যবহৃত হয় **মার্কভ র‍্যান্ডম ফিল্ডস (MRF)**। ১৯৭১ সালের বিখ্যাত **হ্যামার্সলে-ক্লিফোর্ড উপপাদ্য (Hammersley-Clifford Theorem)** প্রমাণ করে যে একটি পজিটিভ ডিস্ট্রিবিউশন কেবল তখনই গ্রাফের শর্তাধীন স্বাধীনতা মেনে চলবে যদি তাকে গ্রাফের ক্লিকের (Cliques) সম্ভাবনার গুণফল বা গিবস ডিস্ট্রিবিউশনে প্রকাশ করা যায়:

$$P(X) = frac1Zprod_Cpsi_C(X_C) = frac1Zexpleft(-sum_CE(X_C)right)$$

এখানে Z হলো নরমালাইজেশন পার্টিশন ফাংশন।

৪. পার্টিশন ফাংশন ও ইনফারেন্সের সংকট

সাধারণ গ্রাফে পার্টিশন ফাংশন Z গণনা করা একটি **#P-complete** চরম জটিল সমস্যা। ফলে জটিল লুপযুক্ত নেটওয়ার্কে নিখুঁত সম্ভাব্যতা বের করা অসম্ভব হয়ে দাঁড়ায়। এর ফলেই জন্ম নেয় ভ্যারিয়েশনাল বায়েস এবং মারকভ চেইন মন্টে কার্লো (MCMC)-এর মতো আধুনিক আসন্ন সমাধান পদ্ধতি।

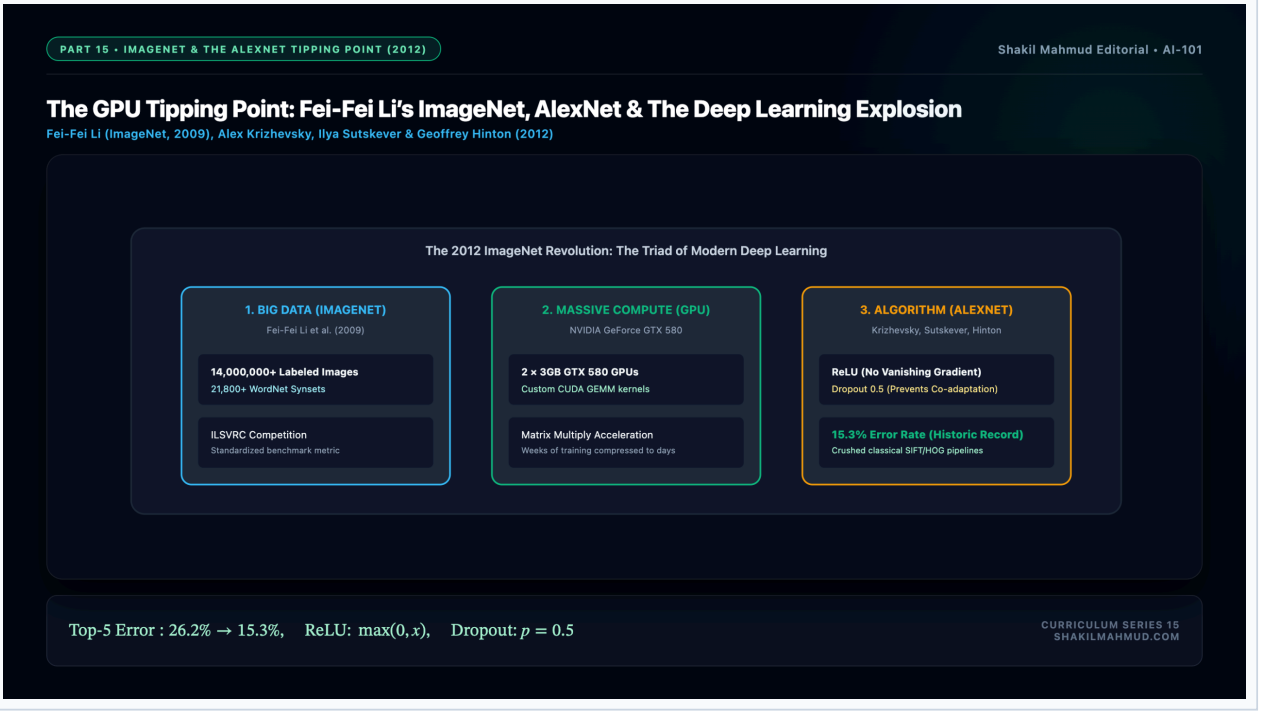
সাইটেশন: মাহমুদ, শাকিল (2026)। "সম্ভাবনাময় গ্রাফিক্যাল মডেল — বায়েসিয়ান নেটওয়ার্ক, এমআরএফ এবং ডি-সেপারেশন"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১৪। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-14-probabilistic-graphical-models-d-separation-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • পঞ্চদশ খণ্ড

লুকায়িত ভেরিয়েবল মডেল — জেনসেনের উপপাদ্য এবং ইএম অ্যালগরিদম

লুকায়িত ভেরিয়েবল মডেল — জেনসেনের উপপাদ্য এবং ইএম অ্যালগরিদম

413 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১৫: *২০১২ সালের যুগান্তকারী ডিপ লার্নিং বিস্ফোরণ: ফেই-ফেই লির ইমেজনেট ডেটাসেট, জিপিইউ প্যারালালেলিজম এবং অ্যালেক্সনেটের রিলু ও ড্রপআউট।*

বাস্তব ডেটা বিশ্লেষণে বহু মৌলিক কারণ থাকে যা কখনো সরাসরি পরিমাপ করা যায় না: যেমন জিনোমিক্সে রোগের অপ্রকাশিত সাব-টাইপ বা টেক্সটের অন্তর্নিহিত টপিক। যখন ডেটাসেটে এমন লুকায়িত চলক Z উপস্থিত থাকে, তখন সামগ্রিক লাইকলিহুড সমীকরণে সামেশনের ওপর লগারিদম বসে যায় —যার ফলে সর্বোচ্চ লাইকলিহুডের স্বাভাবিক সমীকরণগুলো সম্পূর্ণরূপে অকার্যকর হয়ে পড়ে। ১৯৭৭ সালে আর্থার ডেম্পস্টার, ন্যান লেয়ার্ড এবং ডোনাল্ড রুবিন এই সংকটের অমর সমাধান উপহার দেন: **এক্সপেক্টেশন-ম্যাক্সিমাইজেশন (EM) অ্যালগরিদম**।

১. লুকায়িত ভেরিয়েবলের লাইকলিহুড সংকট

ধরা যাক দৃশ্যমান ডেটা X এবং লুকায়িত চলক Z। সামগ্রিক ডেটা পাওয়ার জন্য আমাদের Z-এর সকল মানের ওপর সামেশন করতে হয়:

$$ell(theta) = lnp(Xmidtheta) = sum_{i=1}^{n} lnleft(sum_{z_i} p(x_i, z_i midtheta) right)$$

এখানেই তৈরি হয় মূল সংকট: লগারিদমের ভেতরে সামেশন অবস্থান করায় লগ সরাসরি গুণফলের মতো বিভক্ত হতে পারে না, ফলে প্যারামিটারগুলোর কোনো সহজ সমাধান বের করা যায় না।

Shakil Mahmud • Research Monograph

২. জেনসেনের উপপাদ্য ও এভিডেন্স লোয়ার বাউন্ড (ELBO)

এই জটিলতা কাটাতে আমরা Z-এর ওপর একটি বিকল্প ডিস্ট্রিবিউশন $q(Z)$ কল্পনা করি। সামেশনের ভেতর $q(Z)$ দিয়ে গুণ ও ভাগ করে এবং লগারিদমের অবতল (Concave) প্রকৃতির ওপর **জেনসেনের উপপাদ্য (Jensen's Inequality)** প্রয়োগ করে পাই:

$$lnp(Xmidtheta)gesum_Zq(Z)lnfracp(X,Zmidtheta)q(Z)equivmathcalL(q,theta)$$

এই $mathcalL(q,theta)$-কে বলা হয় **এভিডেন্স লোয়ার বাউন্ড বা ইএলবিও (ELBO)**। এটি প্রমাণ করে যে আমরা সরাসরি আসল লাইকলিহুড স্পর্শ করতে না পারলেও সর্বদা একটি নিরাপদ নিচের সীমানা বা ফ্লোর অপ্টিমাইজ করতে পারি।

৩. ইএম অ্যালগরিদমের দুটি পর্যায়ক্রমিক ধাপ

ইএম অ্যালগরিদম প্রতি পদক্ষেপে দুটি সহজ ধাপে কাজ সম্পন্ন করে:

1. **ই-স্টেপ (E-Step — Expectation):** বর্তমান প্যারামিটার $theta^(t)$ স্থির রেখে $q(Z)$-কে এমনভাবে সাজানো যাতে ইএলবিও আসল লাইকলিহুডের সমান উচ্চতায় পৌঁছায়। এটি ঘটে যখন $q(Z)$ ঠিক বর্তমান পোস্টেরিয়র $p(ZmidX,theta^(t))$-এর সমান হয়। এর মাধ্যমে আমরা প্রত্যাশিত লগ-লাইকলিহুড ফাংশন $Q(theta;theta^(t))$ তৈরি করি।
2. **এম-স্টেপ (M-Step — Maximization):** এই Q-ফাংশনকে প্যারামিটার $theta$-এর সাপেক্ষে সর্বোচ্চ করা হয়:

$$theta^(t+1)=argmax_thetaQ(theta;theta^(t))$$

   যেহেতু এবার লগ সরাসরি জয়েন্ট ডিস্ট্রিবিউশনের ওপর অবস্থান করছে, তাই এর ডেরিভেটিভ বের করে সহজেই নতুন প্যারামিটার পেয়ে যাওয়া যায়।

৪. একমুখী উন্নতির গাণিতিক নিশ্চয়তা

ডেম্পস্টার প্রমুখ প্রমাণ করেন যে ইএম অ্যালগরিদম প্রতি পদক্ষেপে আসল লাইকলিহুডকে **মনোটনিকভাবে বৃদ্ধি করে** ($ell(theta^(t+1))geell(theta^(t))$)। এটি কখনো পেছনের দিকে যায় না এবং নিশ্চিতভাবে একটি লোকাল ম্যাক্সিমামে গিয়ে স্থির হয়।

৫. গাউসিয়ান মিক্সচার মডেল (GMM)

ক্লাস্টারিংয়ের ক্ষেত্রে জিএমএম-এ ইএম অ্যালগরিদম ব্যবহৃত হয়:

- **ই-স্টেপ:** প্রতিটি ডেটা পয়েন্ট কোন ক্লাস্টারে যাওয়ার সম্ভাবনা কত (Responsibility $gamma_ik$) তা বের করা।
- **এম-স্টেপ:** এই সম্ভাবনার ওপর ভিত্তি করে ক্লাস্টারের নতুন কেন্দ্র mu_k, বিস্তার $Sigma_k$ এবং অনুপাত pi_k পুনরায় হিসাব করা।

লুকায়িত রহস্য উন্মোচনে ইএম অ্যালগরিদম আজ আধুনিক পরিসংখ্যানিক কম্পিউটিংয়ের অন্যতম শ্রেষ্ঠ নিদর্শন।

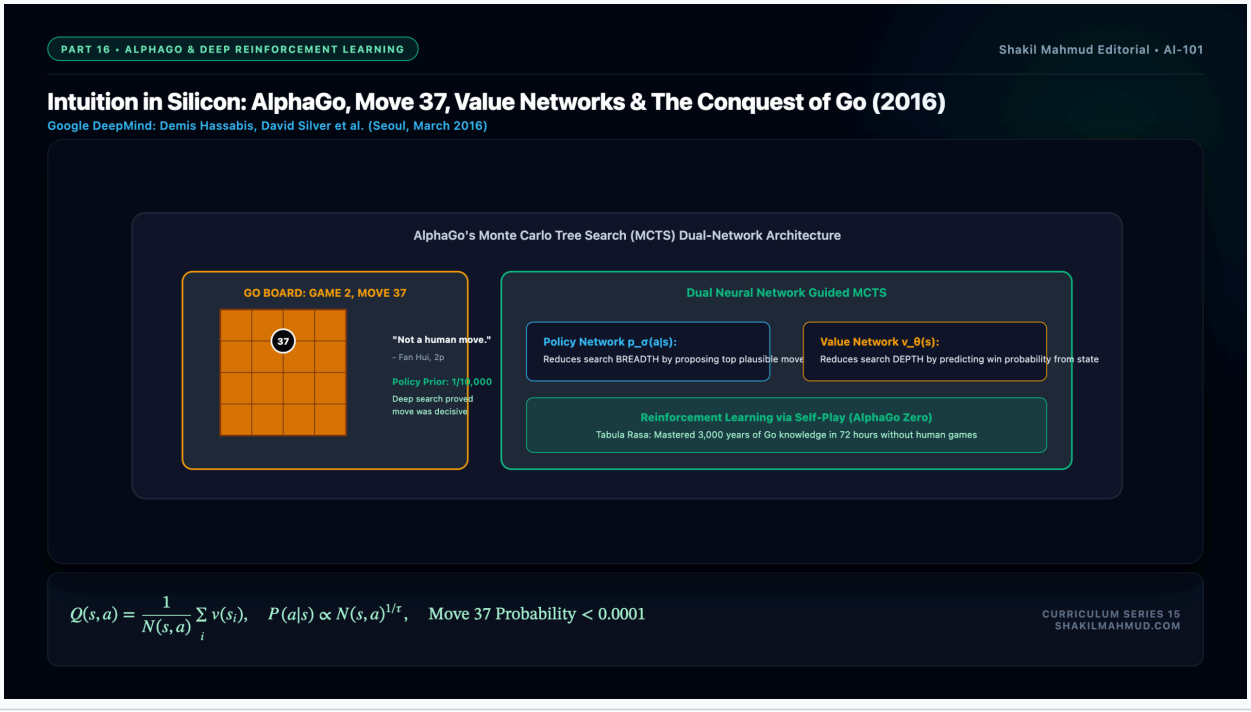
সাইটেশন: মাহমুদ, শাকিল (2026)। "লুকায়িত ভেরিয়েবল মডেল — জেনসেনের উপপাদ্য এবং ইএম অ্যালগরিদম"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১৫।
https://shakilmahmud.com/bn/foundations-of-machine-learning-part-15-latent-variable-models-em-algorithm-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • ষোড়শ খণ্ড

আসন্ন অনুমান ও ভ্যারিয়েশনাল বায়েস — মিন-ফিল্ড রূপরেখা, কেএল ডাইভারজেন্স এবং ইএলবিও

আসন্ন অনুমান ও ভ্যারিয়েশনাল বায়েস — মিন-ফিল্ড রূপরেখা, কেএল ডাইভারজেন্স এবং ইএলবিও

395 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১৬: *আলফাগোর পলিসি ও ভ্যালু নেটওয়ার্ক চালিত মন্টে কার্লো ট্রি সার্চ এবং ঐতিহাসিক চাল ৩৭ (Move 37) এর কম্পিউটার অন্তর্দৃষ্টি।*

বায়েসিয়ান পরিসংখ্যানে নতুন তথ্য X দেখার পর অতীত বিশ্বাস $p(Z)$ পরিবর্তিত হয়ে রূপ নেয় পোস্টেরিয়র ডিস্ট্রিবিউশনে: $p(ZmidX) = fracp(X,Z)p(X)$। কিন্তু হরের মার্জিনাল ইন্টিগ্রাল $p(X) = intp(X,Z)dZ$ গণনা করা উচ্চ মাত্রার স্পেসে গাণিতিকভাবে সম্পূর্ণ অসম্ভব। অতীতে মারকভ চেইন মন্টে কার্লো (MCMC) দিয়ে দৈব স্যাম্পলিং করা হতো, কিন্তু তা ছিল অত্যন্ত ধীরগতির। **ভ্যারিয়েশনাল ইনফারেন্স (Variational Inference — VI)** এই অসম্ভব ইন্টিগ্রেশনকে একটি সাধারণ অপ্টিমাইজেশন বা অবমূল্যায়ন সমস্যায় রূপান্তর করে এক অবিশ্বাস্য বিপ্লব সাধন করে।

১. ভ্যারিয়েশনাল ইনফারেন্সের মূল কাঠামো

আসল জটিল পোস্টেরিয়র $p(ZmidX)$-কে হুবহু নির্ণয় করতে না পেরে আমরা একটি পরিচিত, সহজ ডিস্ট্রিবিউশন পরিবার $mathcalQ = q(Z)$ কল্পনা করি। আমাদের লক্ষ্য হলো এই পরিবারের মধ্য থেকে এমন একটি $q^*(Z)$ খুঁজে বের করা যা আসল পোস্টেরিয়রের সবচেয়ে কাছাকাছি অবস্থান করে। আর দুটি ডিস্ট্রিবিউশনের ব্যবধান পরিমাপের মানদণ্ড হলো **কুলব্যাক-লেইবলার ডাইভারজেন্স (KL Divergence)**:

$$q^*(Z) = argmin_qinmathcalQ D_textKL left(q(Z) parallel p(ZmidX) right)$$

কেএল ডাইভারজেন্সের সংজ্ঞা বিশ্লেষণ করে এবং বায়েসের সূত্র প্রয়োগ করে আমরা পাই আধুনিক পরিসংখ্যানের সবচেয়ে বিখ্যাত পরিচয়:

Shakil Mahmud • Research Monograph

$$lnp(X) = textELBO(q) + D_{textKL}(q(Z)parallelp(ZmidX))$$

২. ইএলবিও (ELBO) বৃদ্ধির মাধ্যমে কেএল হ্রাস

লক্ষ করুন, এই সমীকরণের প্রতিটি অংশ এক গভীর সত্য প্রকাশ করে:

- বামপাশের মোট প্রমাণ $lnp(X)$ হলো q-এর সাপেক্ষে একটি ধ্রুবক।
- যেহেতু কেএল ডাইভারজেন্স কখনোই ঋণাত্মক হতে পারে না ($D_{textKL}ge0$), তাই $textELBO(q)$ সর্বদা আসল লাইকলিহুডের একটি নিচের সীমা (Lower Bound)।
- অতএব, $textELBO(q)$**-কে সর্বোচ্চ করা মানেই হলো স্বয়ংক্রিয়ভাবে কেএল ডাইভারজেন্সকে শূন্যের কাছাকাছি নামিয়ে আনা!**

সবচেয়ে বড় সুবিধা হলো—ইএলবিও গণনা করতে কেবল জয়েন্ট প্রবাবিলিটি $p(X, Z)$ প্রয়োজন হয়, অসম্ভব হর $p(X)$-এর বিন্দুমাত্র দরকার পড়ে না।

৩. মিন-ফিল্ড (Mean-Field) অনুমান

হিসাব সহজ করার জন্য আমরা পদার্থবিজ্ঞান থেকে **মিন-ফিল্ড অনুমান** গ্রহণ করি—যেখানে ধরে নেওয়া হয় সকল লুকায়িত ভেরিয়েবল নিজেদের মধ্যে স্বাধীনভাবে কাজ করে:

$$q(Z) = prod_{j=1}^{M} q_j(Z_j)$$

৪. কোঅর্ডিনেট অ্যাসেন্ট ভ্যারিয়েশনাল ইনফারেন্স (CAVI)

একটি নির্দিষ্ট উপাদান q_j-কে সর্বোচ্চ করতে বাকিগুলোকে স্থির রাখলে এক অপূর্ব প্রতিসাম্য সমীকরণ পাওয়া যায়:

$$q_j^*(Z_j)proptoexpleft(mathbbE_{q_{-j}}left[lnp(X, Z)right]right)$$

অ্যালগরিদমটি একের পর এক ভেরিয়েবলের ওপর এই সমীকরণ প্রয়োগ করে ঘুরতে থাকে যতক্ষণ না সামগ্রিক ইএলবিও সর্বোচ্চ মানে পৌঁছায়। এটি গিবস স্যাম্পলিংয়ের মতো হলেও কোনো দৈব ট্রায়াল ছাড়া সম্পূর্ণ ডিটারমিনিস্টিক উপায়ে কয়েক গুণ দ্রুত কাজ সম্পন্ন করে।

৫. আধুনিক এআই-তে প্রভাব

পরবর্তীতে ২০১৩ সালে ডিডেরিক কিংমা এবং ম্যাক্স ওয়েলিং এই ভ্যারিয়েশনাল সূত্রকেই ডিপ নিউরাল নেটওয়ার্কের সাথে যুক্ত করে আবিষ্কার করেন **ভ্যারিয়েশনাল অটোএনকোডার (VAE)**। স্ট্যাটিস্টিক্যাল ফিজিক্স ও বায়েসিয়ান গণিতের এই মেলবন্ধন আজও জেনারেটিভ এআই বিপ্লবের অন্যতম কেন্দ্রীয় চালিকাশক্তি।

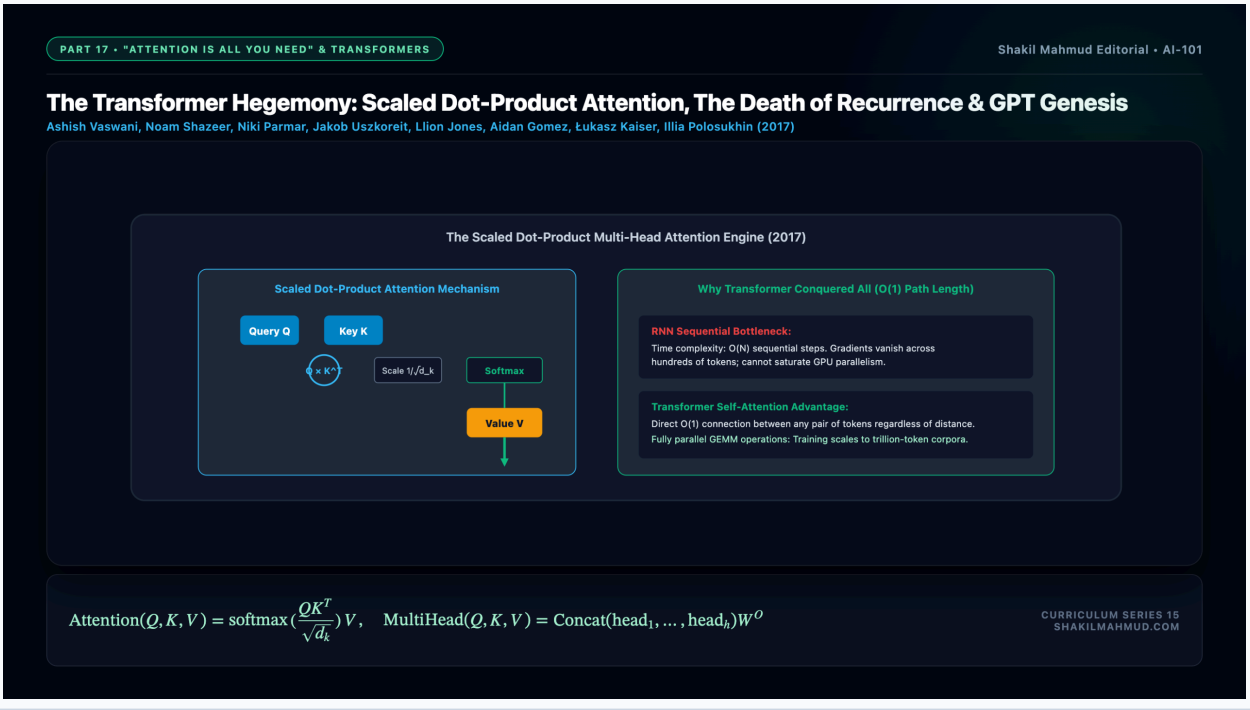
সাইটেশন: মাহমুদ, শাকিল (2026)। "আসন্ন অনুমান ও ভ্যারিয়েশনাল বায়েস — মিন-ফিল্ড রূপরেখা, কেএল ডাইভারজেন্স এবং ইএলবিও"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১৬। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-16-approximate-inference-variational-bayes-elbo-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • সপ্তদশ খণ্ড

গাউসিয়ান প্রসেস ও বায়েসিয়ান নন-প্যারামেট্রিক তত্ত্ব — অসীম মাত্রার ফাংশন স্পেস

গাউসিয়ান প্রসেস ও বায়েসিয়ান নন-প্যারামেট্রিক তত্ত্ব — অসীম মাত্রার ফাংশন স্পেস

381 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১৭: *ট্রান্সফরমার আর্কিটেকচার: স্কেলড ডট-প্রোডাক্ট মাল্টি-হেড সেলফ-অ্যাটেনশন, যা রিকারেন্ট বাধা দূর করে ও(১) টোকেন সংযোগ স্থাপন করেছে।*

চিরায়ত প্যারামেট্রিক মডেলে প্যারামিটার ভেক্টর $thetainmathbbR^d$ একটি নির্দিষ্ট সসীম মাত্রায় আবদ্ধ থাকে, যার ফলে নমুনার সংখ্যা $ntoinfty$ হলেও মডেলের প্রকাশক্ষমতা সীমাবদ্ধ থাকে। **বায়েসিয়ান নন-প্যারামেট্রিক (Bayesian Non-Parametrics)** তত্ত্ব এই সীমাবদ্ধতা ভেঙে দেয়। এই ধারার সবচেয়ে মার্জিত কাঠামো হলো **গাউসিয়ান প্রসেস (Gaussian Process – GP)**—যা কোনো নির্দিষ্ট ভেক্টরের ওপর নয়, বরং সরাসরি অসীম মাত্রার অবিচ্ছিন্ন ফাংশনাল স্পেস $mathcalF = f : mathcalXtomathbbR$-এর ওপর সম্ভাব্যতা বণ্টন আরোপ করে।

১. গাণিতিক সংজ্ঞা এবং ধারাবাহিকতা

গাণিতিকভাবে, একটি গাউসিয়ান প্রসেস হলো অগণিত র‍্যান্ডম ভ্যারিয়েবলের একটি সমন্বয়, যার যেকোনো সসীম উপসেটের যৌথ বিন্যাস সর্বদা একটি বহু-চলকবিশিষ্ট গাউসিয়ান বিন্যাস (Multivariate Gaussian)। একটি প্রসেস $f(x)$ তার গড় ফাংশন $m(x)$ এবং ধনাত্মক আধা-নির্দিষ্ট কোভ্যারিয়েন্স কার্নেল k(x, x’) দ্বারা সম্পূর্ণ নির্ধারিত হয়:

f(x) sim mathcal{GP}left(m(x), k(x, x’) right)

যেখানে কলমোগোরোভ এক্সটেনশন উপপাদ্য (Kolmogorov Extension Theorem) নিশ্চিত করে যে এই অসীম মাত্রার স্টোক্যাস্টিক প্রক্রিয়া গাণিতিকভাবে সম্পূর্ণরূপে বৈধ ও সামঞ্জস্যপূর্ণ।

Shakil Mahmud • Research Monograph

২. কোভ্যারিয়েন্স কার্নেল এবং মসৃণতা

নমুনা ফাংশনের মসৃণতা ও আচরণ নির্ধারিত হয় কার্নেল k(x, x’)-এর মাধ্যমে:

- **স্কয়ার্ড এক্সপোনেনশিয়াল (RBF কার্নেল):** $k_{textSE}(r) = sigma_f^2 expleft(-r^2/2ell^2right)$। এটি অসীম অন্তরীকরণযোগ্য (C^infty) অতি-মসৃণ ফাংশন তৈরি করে।
- **ম্যাটার্ন কার্নেল (Matérn Class):** যখন $nu = 3/2$ বা $5/2$, ফাংশনগুলো বাস্তব জগতের তরঙ্গ ও সংকেতের মতো বাস্তবিক মাত্রার অমসৃণতা ধারণ করতে পারে।
- **পর্যায়বৃত্ত কার্নেল (Periodic Kernel):** নির্দিষ্ট পর্যায়কাল p অনুসারে আবর্তিত চক্রাকার প্যাটার্ন মডেল করতে ব্যবহৃত হয়।

৩. নিখুঁত অ্যানালিটিক্যাল পোস্টেরিয়র ইনফারেন্স

শব্দযুক্ত পর্যবেক্ষণ $y = f(x) + epsilon$ বিবেচনা করি যেখানে $epsilonsimmathcal N(0, sigma_n^2)$। পর্যবেক্ষণ বিন্দু X এবং নতুন কোয়েরি X_*-এর জন্য যৌথ গাউসিয়ান ডিস্ট্রিবিউশন শর্তাধীন করার মাধ্যমে আমরা নিখুঁত পোস্টেরিয়র সমাধান পাই:

$$mu_* = K(X_*, X)left[K(X, X) + sigma_n^2 Iright]^{-1} y$$

Sigma_* = K(X_*, X_*) – K(X_*, X) left[K(X, X) + sigma_n^2 I right]^{-1} K(X, X_*)

এই সমীকরণের সবচেয়ে চমকপ্রদ বৈশিষ্ট্য হলো এর অনিশ্চয়তা পরিমাপ ($Sigma_*$): পরিচিত ডেটার সন্নিকটে ভ্যারিয়েন্স সংকুচিত হয়, কিন্তু অজানা এলাকায় তা প্রসারিত হয়ে মডেলের অজ্ঞতা নিখুঁতভাবে প্রকাশ করে।

৪. মার্জিনাল লাইকলিহুড এবং বায়েসিয়ান অপ্টিমাইজেশন

মডেলের হাইপারপ্যারামিটার নির্ধারণে এমপিরিক্যাল বায়েস পদ্ধতিতে লগ মার্জিনাল লাইকলিহুড সর্বোচ্চকরণ করা হয়:

log p(y mid X, theta) = -frac{1}{2} y^T (K_y)^{-1} y – frac{1}{2} log |K_y| – frac{n}{2}log(2pi)

এখানে দ্বিতীয় পদ $-frac12log|K_y|$ একটি প্রাকৃতিক ওকামের ক্ষুর (Occam’s Razor) হিসেবে কাজ করে যা জটিল মডেলকে স্বয়ংক্রিয়ভাবে শাস্তি দেয়। গাউসিয়ান প্রসেস ব্যয়বহুল ব্ল্যাক-বক্স অপ্টিমাইজেশনে (Bayesian Optimization) একুইজিশন ফাংশন (যেমন GP-UCB) ব্যবহারের মাধ্যমে সর্বোত্তম অন্বেষণ নিশ্চিত করে।

সাইটেশন: মাহমুদ, শাকিল (2026)। "গাউসিয়ান প্রসেস ও বায়েসিয়ান নন-প্যারামেট্রিক তত্ত্ব — অসীম মাত্রার ফাংশন স্পেস"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১৭।
https://shakilmahmud.com/bn/foundations-of-machine-learning-part-17-gaussian-processes-bayesian-nonparametrics-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • অষ্টাদশ খণ্ড

ডাইমেনশনালিটি রিডাকশন ও ম্যানিফোল্ড লার্নিং — পিসিএ থেকে ইউম্যাপ

ডাইমেনশনালিটি রিডাকশন ও ম্যানিফোল্ড লার্নিং — পিসিএ থেকে ইউম্যাপ

389 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১৮: *নিউরাল স্কেলিং সূত্র: কাপলান ও চিনচিলা পাওয়ার-ল সীমান্ত এবং জিপিটি-৩ তে ইন-কনটেক্সট ফিউ-শট প্রম্পটিংয়ের আত্মপ্রকাশ।*

উচ্চ-মাত্রিক ডেটা স্পেসে কাজ করার সময় রিচার্ড বেলম্যানের কুখ্যাত *কার্স অফ ডাইমেনশনালিটি (Curse of Dimensionality)* প্রকট হয়ে ওঠে: মাত্রা $dtoinfty$ হলে মহাশূন্যের আয়তন এক্সপোনেনশিয়ালি বৃদ্ধি পায় এবং সমস্ত ডেটা পয়েন্ট একে অপরের থেকে সমদূরবর্তী ও বিচ্ছিন্ন হয়ে পড়ে। তবে বাস্তব জগতের ডেটা সর্বত্র সমভাবে ছড়ানো থাকে না; এটি সাধারণত একটি নিম্ন-মাত্রিক নন-লিনিয়ার সাব-ম্যানিফোল্ড $mathcalMsubsetmathbbR^d$-এর পৃষ্ঠে অবস্থান করে যার অন্তর্নিহিত মাত্রা $d*lld$। **ডাইমেনশনালিটি রিডাকশন (Dimensionality Reduction)** এই অন্তর্নিহিত জ্যামিতি অক্ষুণ্ণ রেখে ডেটাকে নিম্ন মাত্রায় রূপান্তর করে।

১. প্রিন্সিপাল কম্পোনেন্ট অ্যানালাইসিস (PCA)

১৯০১ সালে কার্ল পিয়ারসন **পিসিএ (PCA)** প্রবর্তন করেন। কেন্দ্রিকৃত ডেটা ম্যাট্রিক্স $XinmathbbR^{ntimesd}$-এর জন্য পিসিএ এমন কিছু অর্থোগোনাল একক ভেক্টর w খুঁজে বের করে যা প্রক্ষিপ্ত উপাত্তের ভ্যারিয়েন্স সর্বোচ্চ করে:

$$max_{w:|w|=1} w^T SigmawimpliesSigmaw = lambdaw$$

ল্যাগ্রাঞ্জ মাল্টিপ্লায়ার প্রয়োগের মাধ্যমে কোভ্যারিয়েন্স ম্যাট্রিক্স $Sigma$-এর শীর্ষ আইগেনভ্যালু সম্পর্কিত আইগেনভেক্টরগুলোই প্রধান উপাংশ হিসেবে চিহ্নিত হয়। একার্ট-ইয়ং-মিরস্কি উপপাদ্য অনুসারে ট্রাঙ্কেটেড এসভিডি (SVD) ফ্রবেনিয়াস নর্মে ন্যূনতম ত্রুটিযুক্ত সর্বোত্তম লিনিয়ার প্রক্ষেপণ নিশ্চিত করে।

Shakil Mahmud • Research Monograph

২. নন-লিনিয়ার স্পেকট্রাল এম্বেডিং: কার্নেল পিসিএ এবং আইসোম্যাপ

ম্যানিফোল্ড যখন অ-রৈখিকভাবে বাঁকা থাকে (যেমন সুইস রোল), তখন লিনিয়ার পিসিএ ব্যর্থ হয়। ১৯৯৮ সালে বার্নহার্ড শোলকফ **কার্নেল পিসিএ (Kernel PCA)** উদ্ভাবন করেন যা ডেটাকে উচ্চ-মাত্রিক আরকেএইচএস-এ প্রক্ষেপ করে নন-লিনিয়ার প্রধান উপাংশ নির্ণয় করে।

২০০০ সালে টেনেবাম প্রবর্তন করেন **আইসোম্যাপ (Isomap)**। এটি ইউক্লিডীয় দূরত্বের বদলে গ্রাফের মাধ্যমে **জিওডেসিক দূরত্ব (Geodesic Distance)** পরিমাপ করে এবং ক্লাসিক্যাল মাল্টি-ডাইমেনশনাল স্কেলিং (MDS) প্রয়োগ করে বাঁকানো ম্যানিফোল্ডকে সমতলে নিখুঁতভাবে উন্মুক্ত করে।

৩. ল্যাপ্লাসিয়ান আইগেনম্যাপস (Laplacian Eigenmaps)

বেলকিন এবং নিয়গি (২০০৩) স্পেকট্রাল গ্রাফ থিওরির ওপর ভিত্তি করে ল্যাপ্লাসিয়ান আইগেনম্যাপস তৈরি করেন। গ্রাফ ল্যাপ্লাসিয়ান L = D – W অপ্টিমাইজেশনের মাধ্যমে স্থানীয় নৈকট্য রক্ষা পায়:

$$min_Y textTr(Y^T LY) quad texts.t. quad Y^T DY = I$$

আইগেনসিস্টেম $Lv = lambda Dv$-এর ক্ষুদ্রতম ধনাত্মক আইগেনভেক্টরগুলো রিম্যানিয়ান ম্যানিফোল্ডের ল্যাপ্লাস-বেল্ট্রামি অপারেটরের সাথে সরাসরি সম্পর্কিত।

৪. t-SNE এবং UMAP: আধুনিক ক্লাস্টার ভিজ্যুয়ালাইজেশন

২০০৮ সালে লরেন্স ভ্যান ডার মাটেন এবং জিওফ্রে হিন্টন উদ্ভাবন করেন **t-SNE**। উচ্চ মাত্রায় বিন্দুসমূহের দূরত্ব গাউসিয়ান ডিস্ট্রিবিউশন দিয়ে সম্ভাবনায় রূপান্তরিত হয় এবং নিম্ন মাত্রায় স্টুডেন্টস-টি ডিস্ট্রিবিউশন (Student-t with 1 DOF) ব্যবহার করা হয়। এটি বিখ্যাত *ক্রাউডিং সমস্যা (Crowding Problem)* দূর করে নিখুঁত ক্লাস্টার উন্মোচন করে।

২০১৮ সালে লিল্যান্ড ম্যাকইনস **UMAP** উদ্ভাবন করেন, যা ক্যাটাগরি থিওরি এবং রিম্যানিয়ান জ্যামিতির ওপর প্রতিষ্ঠিত। এটি অস্পষ্ট সিম্প্লিসিয়াল সেট (Fuzzy Simplicial Sets) বিশ্লেষণের মাধ্যমে t-SNE-এর চেয়ে বহুগুণ দ্রুত কাজ করে এবং বৈশ্বিক কাঠামোগত সম্পর্ক নিখুঁতভাবে সংরক্ষণ করে।

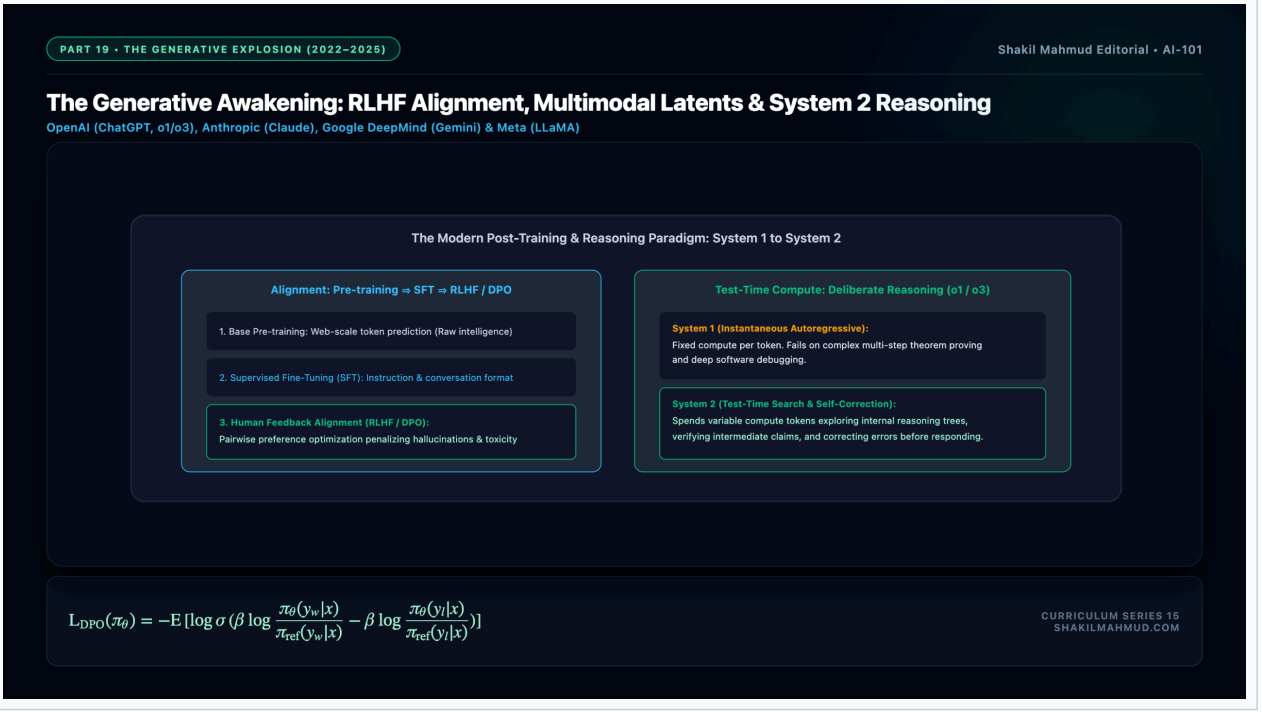
সাইটেশন: মাহমুদ, শাকিল (2026)। "ডাইমেনশনালিটি রিডাকশন ও ম্যানিফোল্ড লার্নিং — পিসিএ থেকে ইউম্যাপ"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১৮।
https://shakilmahmud.com/bn/foundations-of-machine-learning-part-18-dimensionality-reduction-spectral-manifold-learning-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • ঊনবিংশ খণ্ড

সেমি-সুপারভাইজড ও অ্যাক্টিভ লার্নিং — স্বল্প-লেবেল রেজিমের ইনফরমেশন জিওমেট্রি

সেমি-সুপারভাইজড ও অ্যাক্টিভ লার্নিং — স্বল্প-লেবেল রেজিমের ইনফরমেশন জিওমেট্রি

369 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ১৯: *আধুনিক জেনারেটিভ এআই বিপ্লব: আরএলএইচএফ ও ডিপিও অ্যালাইনমেন্ট থেকে সিস্টেম-২ টেস্ট-টাইম রিজনিং সার্চের রূপরেখা।*

সুপারভাইজড লার্নিং সম্পূর্ণভাবে বিশাল লেবেলযুক্ত ডেটাসেট $mathcal{D}_L = (x_i, y_i)_{i=1}^{l}$-এর ওপর নির্ভরশীল। তবে চিকিৎসাবিজ্ঞান, কোয়ান্টাম রসায়ন বা বায়োইনফরমেটিক্সের মতো সংবেদনশীল ক্ষেত্রে ডেটায় লেবেল যুক্ত করা অত্যন্ত ব্যয়বহুল ও সময়সাপেক্ষ। বিপরীতে লেবেলবিহীন কাঁচা ডেটা $mathcal{D}_U$ প্রচুর পরিমাণে বিদ্যমান $(uggl)$। **সেমি-সুপারভাইজড লার্নিং (SSL)** এবং **অ্যাক্টিভ লার্নিং (AL)** লেবেলবিহীন তথ্যের অন্তর্নিহিত জ্যামিতি ব্যবহার করে ন্যূনতম খরচে মডেলের কার্যকারিতা বহুগুণ বৃদ্ধি করে।

১. সেমি-সুপারভাইজড শিখনের তিনটি মৌলিক ভিত্তি

লেবেলবিহীন ডেটা $P(X)$ টার্গেট প্রেডিকশন $P(Y midX)$-এ তখনই সহায়তা করতে পারে যখন তাদের মাঝে কাঠামোগত সম্পর্ক থাকে। এটি তিনটি ভিত্তির ওপর দাঁড়িয়ে আছে:

1. **স্মুথনেস অনুমান (Smoothness Assumption):** ডেটা স্পেসের উচ্চ ঘনত্বের অঞ্চলে যদি দুটি বিন্দু কাছাকাছি থাকে, তবে তাদের লেবেলও সমরূপ হবে।
2. **ক্লাস্টার অনুমান (Cluster Assumption):** ক্লাসিফিকেশনের সীমানা কখনই উচ্চ ঘনত্বের ক্লাস্টারের মধ্য দিয়ে যাবে না; সীমানা সর্বদা ফাঁকা উপত্যকার মধ্য দিয়ে অতিক্রম করবে।
3. **ম্যানিফোল্ড অনুমান (Manifold Assumption):** জটিল ডেটা নিম্ন-মাত্রিক ম্যানিফোল্ডের পৃষ্ঠে আবদ্ধ থাকে এবং সেই পৃষ্ঠ বরাবর পরিবর্তন সুষম হয়।

Shakil Mahmud • Research Monograph

২. গ্রাফ হারমোনিক ফাংশন ও লেবেল প্রপাগেশন

২০০৩ সালে শাওজিন ঝু, জুবিন ঘাহরামানি এবং জন ল্যাফার্টি গ্রাফের ওপর **হারমোনিক ফাংশন (Harmonic Functions)** ব্যবহারের মাধ্যমে সেমি-সুপারভাইজড লার্নিংয়ের নিখুঁত রূপরেখা দেন। গ্রাফের ডিরিকলেট শক্তি (Dirichlet Energy) ল্যাপ্লাসিয়ান ম্যাট্রিক্সের মাধ্যমে প্রকাশ করা হয়:

mathcal{E}(f) = frac{1}{2} sum_{i,j=1}^n W_{ij} (f(x_i) – f(x_j))^2 = f^T L f

পরিচিত লেবেলযুক্ত বিন্দুগুলোকে অপরিবর্তিত রেখে শক্তি সর্বনিম্ন করলে লেবেলবিহীন বিন্দুর জন্য ক্লোজড-ফর্ম সমাধান পাওয়া যায়:

$$f_U^* = -L_{UU}^{-1} L_{UL} y_L$$

এই সমীকরণটি প্রমাণ করে যে এটি একটি অ্যাবজরবিং মার্কভ চেইনের সমতুল্য, যেখানে প্রতিটি লেবেলবিহীন বিন্দুতে মান হলো একটি র‍্যান্ডম ওয়াকারের নির্দিষ্ট ক্লাসে শোষিত হওয়ার সম্ভাবনা।

৩. মাল্টি-ভিউ কো-ট্রেনিং (Co-Training)

১৯৯৮ সালে এভরিম ব্লাম এবং টম মিচেল **কো-ট্রেনিং** প্রবর্তন করেন। যদি কোনো ডেটার দুটি স্বাধীন বৈশিষ্ট্য থাকে (যেমন ওয়েবপৃষ্ঠার টেক্সট এবং বহিরাগত অ্যাঙ্কর লিঙ্ক), তবে দুটি পৃথক মডেল প্রশিক্ষণ দিয়ে তাদের পারস্পরিক মতৈক্যের মাধ্যমে লেবেলবিহীন ডেটা স্বয়ংক্রিয়ভাবে লেবেল করা সম্ভব।

৪. অ্যাক্টিভ লার্নিং ও তথ্য আহরণ

মডেল যখন নিজেই সিদ্ধান্ত নেয় কোন ডেটা পয়েন্টটির লেবেল জানা সবচেয়ে জরুরি, তখন তাকে **অ্যাক্টিভ লার্নিং** বলে। এর প্রধান কৌশলগুলো হলো:

- **অনিশ্চয়তা নমুনায়ন (Uncertainty Sampling):** শ্যানন এন্ট্রপি $H(Ymidx)$ সর্বোচ্চ এমন জটিল পয়েন্টগুলোর লেবেল জানতে চাওয়া।
- **ক্যোয়ারী বাই কমিটি (Query by Committee):** একাধিক ভিন্ন মডেলের সমন্বয়ে গঠিত কমিটির সদস্যদের মতবিরোধ যেখানে সবচেয়ে বেশি, সেই ডেটা পয়েন্ট নির্বাচন করা।
- **ফিশার ইনফরমেশন ম্যাক্সিমাইজেশন:** প্যারামিটারের অনিশ্চয়তা সর্বনিম্ন করতে ফিশার ইনফরমেশন ম্যাট্রিক্স $mathcalI(theta)$ অপ্টিমাইজ করা।

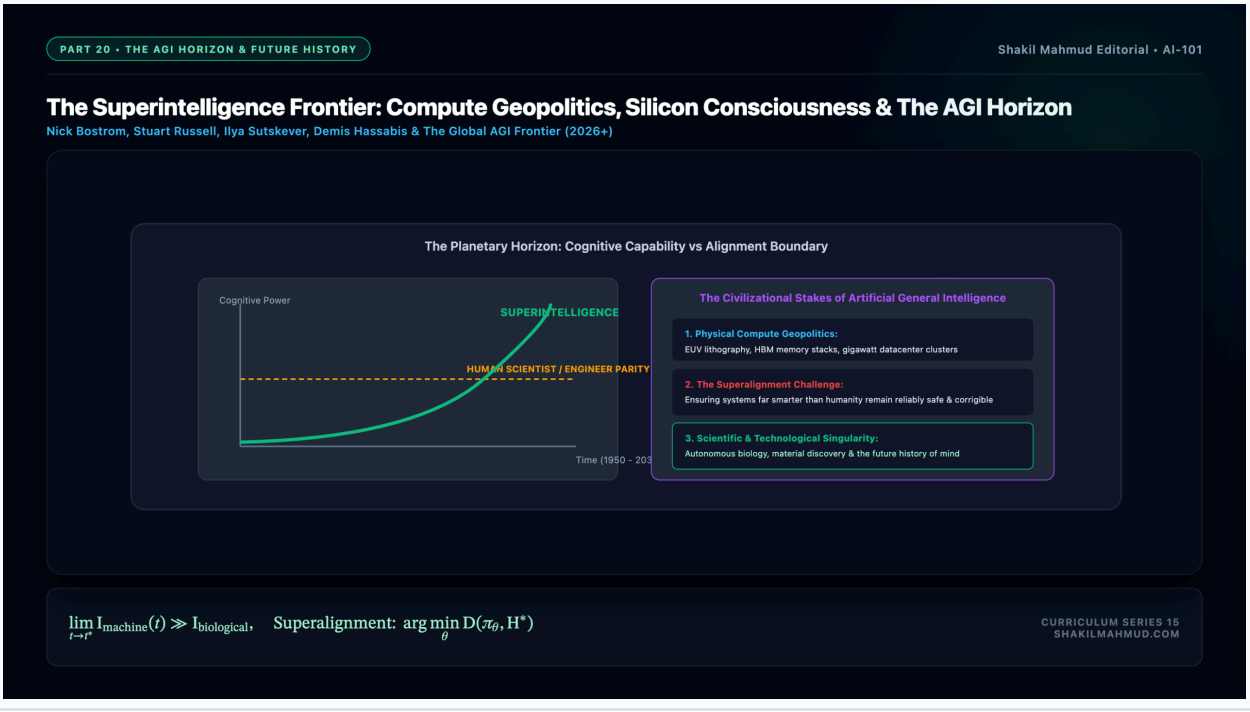
সাইটেশন: মাহমুদ, শাকিল (2026)। "সেমি-সুপারভাইজড ও অ্যাক্টিভ লার্নিং — স্বল্প-লেবেল রেজিমের ইনফরমেশন জিওমেট্রি"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ১৯। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-19-semi-supervised-active-learning-information-geometry-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • বিংশ খণ্ড

মডেল ক্যালিব্রেশন ও কনফরমাল প্রেডিকশন — সসীম নমুনায় নিশ্চিত পূর্বাভাসের গাণিতিক গ্যারান্টি

মডেল ক্যালিব্রেশন ও কনফরমাল প্রেডিকশন — সসীম নমুনায় নিশ্চিত পূর্বাভাসের গাণিতিক গ্যারান্টি

464 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



চিত্র ২০: *কৃত্রিম সাধারণ বুদ্ধিমত্তা (AGI) দিগন্ত: সূচকীয় সক্ষমতা বৃদ্ধি, সেমিকন্ডাক্টর কম্পিউট অবকাঠামো এবং সুপারঅ্যালাইনমেন্টের চ্যালেঞ্জ।*

আধুনিক মেশিন লার্নিং এবং গভীর নিউরাল নেটওয়ার্কগুলো বেঞ্চমার্কে মানুষের চেয়েও বেশি নির্ভুলতা প্রদর্শন করে ঠিকই, কিন্তু তাদের ভেতরে একটি মারাত্মক কাঠামোগত দুর্বলতা লুকিয়ে থাকে: **ভ্রান্ত আত্মবিশ্বাস বা মিসক্যালিব্রেশন (Miscalibration)**। একটি ডিপ লার্নিং মডেল যখন কোনো পূর্বাভাসের পেছনে ৯৯% সম্ভাবনা দাবি করে, অনেক সময় বাস্তবে দেখা যায় তার সত্য হওয়ার হার মাত্র ৮০%। ক্যান্সার শনাক্তকরণ, স্বায়ত্তশাসিত গাড়ি পরিচালনা বা আইনি সিদ্ধান্তের মতো স্পর্শকাতর ক্ষেত্রে অনিশ্চয়তার সঠিক পরিমাপ ছাড়া পূর্বাভাস বিপজ্জনক। **কনফরমাল প্রেডিকশন (Conformal Prediction)** এই সংকটের সমাধান দেয়—এটি কোনো পূর্বানুমান ছাড়াই সসীম নমুনায় গাণিতিকভাবে নিশ্চিত প্রেডিকশন সেট প্রদান করে।

১. মিসক্যালিব্রেশন ও এক্সপেক্টেড ক্যালিব্রেশন এরর (ECE)

একটি মডেলকে নিখুঁতভাবে ক্যালিব্রেটেড বলা হয় যদি তার দাবিকৃত সম্ভাব্যতার মান বাস্তব নির্ভুলতার হারের হুবহু সমান হয়:

$$mathbbPleft(hatY = YmidhatP = pright) = p, quad forallpin[0,1]$$

২০১৭ সালে চুয়ান গুও এবং সহকর্মীরা দেখান যে আধুনিক গভীর নেটওয়ার্কগুলো অতিরিক্ত আত্মবিশ্বাসী (Overconfident)। এই বিচ্যুতি পরিমাপে **Expected Calibration Error (ECE)** ব্যবহার করা হয়:

Shakil Mahmud • Research Monograph

text{ECE} = sum_{m=1}^M frac{|B_m|}{n} left| text{acc}(B_m) – text{conf}(B_m) right|

২. পোস্ট-হক ক্যালিব্রেশন ও টেম্পারেচার স্কেলিং

মডেলের অন্তর্নিহিত পারফরম্যান্স অক্ষুণ্ন রেখে সম্ভাব্যতা সংশোধন করার কৌশলগুলোর মধ্যে প্রধান হলো:

- **প্ল্যাট স্কেলিং (Platt Scaling):** ১৯৯৯ সালে জন প্ল্যাট এসভিএম মার্জিনে লজিস্টিক সিগময়েড ফাংশন প্রয়োগের মাধ্যমে সম্ভাব্যতা সংশোধন করেন।
- **আইসোটোনিক রিগ্রেশন (Isotonic Regression):** মোনোটোনিক স্টেপ ফাংশন ব্যবহারের মাধ্যমে একটি নন-প্যারামেট্রিক ক্যালিব্রেশন পদ্ধতি।
- **টেম্পারেচার স্কেলিং (Temperature Scaling):** ডিপ নিউরাল নেটওয়ার্কের লগিট আউটপুটগুলোকে একটি মাত্র ধনাত্মক স্কেলার T > 0 দিয়ে ভাগ করা হয়:

$$hatp_k(x) = fracexp(z_k(x)/T)sum_{j=1}^{K}exp(z_j(x)/T)$$

যখন T > 1, পূর্বাভাসিত ক্লাসের ক্রম পরিবর্তন না করেই মডেলের অতিরিক্ত অহংকার ও ওভারকনফিডেন্স প্রশমিত হয়।

৩. কনফরমাল প্রেডিকশন: বণ্টন-নিরপেক্ষ গাণিতিক নিশ্চয়তা

১৯৯৯ সালে ভ্লাদিমির ভভক এবং আলেকজান্ডার গ্যামারম্যান **কনফরমাল প্রেডিকশন** তত্ত্ব আবিষ্কার করেন। এটি কোনো একক বিন্দুর অনিশ্চিত পূর্বাভাসের বদলে একটি সেট $C(X)subsetmathcalY$ তৈরি করে, যা ব্যবহারকারীর নির্ধারিত আত্মবিশ্বাস স্তর 1 – alpha-এ গাণিতিকভাবে নিখুঁত কভারেজ নিশ্চিত করে:

mathbb{P}left(Y_{n+1} in C(X_{n+1}) right) ge 1 – alpha

সবচেয়ে গুরুত্বপূর্ণ বৈশিষ্ট্য হলো—এই নিশ্চয়তা **বণ্টন-নিরপেক্ষ (Distribution-Free)**। ডেটার কোনো অজানা জটিল বণ্টন থাকলেও বা যেকোনো ব্ল্যাক-বক্স মডেল ব্যবহার করা হলেও এটি সমানভাবে কার্যকর, যদি কেবল ডেটা পয়েন্টগুলো *এক্সচেঞ্জেবল (Exchangeable)* হয়।

৪. স্প্লিট কনফরমাল প্রেডিকশন ও স্যান্ডউইচ বাউন্ড

স্প্লিট কনফরমাল পদ্ধতিতে ডেটাকে ট্রেইনিং এবং ক্যালিব্রেশন সেটে বিভক্ত করা হয়:

1. ক্যালিব্রেশন সেটের প্রতিটি বিন্দুর জন্য একটি **নন-কনফরমিটি স্কোর (Non-Conformity Score)** $s_i = s(x_i, y_i)$ নির্ণয় করা হয় (যেমন ক্লাসিফিকেশনে 1 – hat{P}(Y = y mid x))।
2. স্কোরগুলোর এমপিরিক্যাল কোয়ান্টাইল $hatq = textQuantile(s_1, dots, s_n; fraclceil(n+1)(1-alpha)rceiln)$ গণনা করা হয়।
3. নতুন যেকোনো টেস্ট ইনপুটের জন্য প্রেডিকশন সেট গঠিত হয়: $C(X_{n+1}) = yinmathcalY : s(X_{n+1}, y)lehatq$।

এক্সচেঞ্জিবিলিটি নীতির কারণে টেস্ট বিন্দুর স্কোর ক্যালিব্রেশন স্কোরের মাঝে সমভাবে অবস্থান করে, যা নিম্নোক্ত নিখুঁত স্যান্ডউইচ বাউন্ড প্রদান করে:

1 – alpha le mathbb{P}left(Y_{n+1} in C(X_{n+1}) right) le 1 – alpha + frac{1}{n+1}

কনফরমাল প্রেডিকশন আধুনিক কৃত্রিম বুদ্ধিমত্তার অনিশ্চয়তাকে একটি সুনির্দিষ্ট, নির্ভরযোগ্য ও নিয়ন্ত্রণযোগ্য গাণিতিক শৃঙ্খলায় আবদ্ধ করে।

❦

সাইটেশন: মাহমুদ, শাকিল (2026)। "মডেল ক্যালিব্রেশন ও কনফরমাল প্রেডিকশন — সসীম নমুনায় নিশ্চিত পূর্বাভাসের গাণিতিক গ্যারান্টি"। মেশিন লার্নিং ও পরিসংখ্যানিক শিখন তত্ত্ব, খণ্ড ২০। https://shakilmahmud.com/bn/foundations-of-machine-learning-part-20-model-calibration-conformal-prediction-uncertainty-bn/

Shakil Mahmud • Research Monograph

আকর গ্রন্থ ও গবেষণা সূত্র

নির্বাচিত গ্রন্থপঞ্জি

এই মনোগ্রাফে আলোচিত তত্ত্ব, গাণিতিক প্রতিপাদন এবং ঐতিহাসিক বিশ্লেষণের ভিত্তি হিসেবে নিম্নলিখিত আকর গ্রন্থ ও প্রাথমিক গবেষণাপত্রসমূহ ব্যবহৃত হয়েছে:

Vapnik, Vladimir N. (1998). *Statistical Learning Theory*. Wiley-Interscience, New York.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, New York.

Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York.

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.

Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA.

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer Science & Business Media.

লেখক পরিচিতি

লেখক পরিচিতি

❖

Shakil Mahmud

গবেষক, লেখক ও সফটওয়্যার প্রকৌশলী

শাকিল মাহমুদ একজন স্বাধীন গবেষক, লেখক ও সফটওয়্যার প্রকৌশলী। তাঁর প্রধান আগ্রহের বিষয় হলো তাত্ত্বিক পদার্থবিজ্ঞান, গণিতের ইতিহাস, আধুনিক ম্যাক্রোঅর্থনীতি, কম্পিউটিংয়ের বিবর্তন এবং স্থান-কালের দার্শনিক তাৎপর্য। তিনি জটিল দার্শনিক ও বৈজ্ঞানিক ধারণাকে গভীর ও সুললিত ভাষায় উপস্থাপন করতে পছন্দ করেন। তাঁর সার্বিক কাজ পাওয়া যাবে তাঁর ব্যক্তিগত গবেষণা পোর্টাল **shakilmahmud.com** এ।

COLOPHON

Typeset in EB Garamond (English) and Kohinoor Bangla / Noto Serif Bengali (Bengali).
Mathematical typesetting generated with Donald Knuth's TeX algorithms and KaTeX.
Published digitally as a monolithic research monograph by Shakil Mahmud Editorial Series.
Dhaka & Singapore • 2026 • https://shakilmahmud.com