Shakil Mahmud • Research Monograph

SHAKIL MAHMUD EDITORIAL & RESEARCH MONOGRAPHS

ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব

ব্যাকপ্রোপাগেশন ও সার্বজনীন অনুমিতিকরণ থেকে কনভোলিউশন, রেসনেট, ট্রান্সফরমার, সেলফ-অ্যাটেনশন, রোটারি এমবেডিং, ভিএই এবং স্টেট-স্পেস মডেলের ২০ খণ্ডের এক গভীর স্থাপত্যিক মহাকাব্য।

SHAKIL MAHMUD

গবেষক ও লেখক

SHAKIL MAHMUD EDITORIAL & RESEARCH SERIES

প্রথম সংকলিত সংস্করণ • ঢাকা ও সিঙ্গাপুর • ২০২৬

ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব

ব্যাকপ্রোপাগেশন ও সার্বজনীন অনুমিতিকরণ থেকে কনভোলিউশন, রেসনেট, ট্রান্সফরমার, সেলফ-অ্যাটেনশন, রোটারি এমবেডিং, ভিএই এবং স্টেট-স্পেস মডেলের ২০ খণ্ডের এক গভীর স্থাপত্যিক মহাকাব্য।

LIBRARY & CATALOGING-IN-PUBLICATION DATA

Title: ডিপ লার্নিং ও রিপ্রেজেন্টেশন 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/deep-learning/

প্রস্তাবিত সাইটেশন:
Mahmud, Shakil. (2026). *ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব: The Complete 20-Part Collected Monograph*. Shakil Mahmud Research Series. https://shakilmahmud.com/series/deep-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

❖

ডিপ লার্নিং হলো হায়ারারকিক্যাল রিপ্রেজেন্টেশন শিখনের বিজ্ঞান—যা জটিল উচ্চ-মাত্রিক সংকেত থেকে অন্তর্নিহিত বৈশিষ্ট্যসমূহকে উন্মোচন করে। বিশ খণ্ডে সংকলিত এই গবেষণা মনোগ্রাফ গভীর নিউরাল আর্কিটেকচারের গাণিতিক ভিত্তি পুঙ্খানুপুঙ্খভাবে উন্মোচন করে: রিম্যানিয়ান ব্যাকপ্রোপাগেশন, কনভোলিউশনাল সমতুল্যতা, রেসিডুয়াল গ্রেডিয়েন্ট প্রবাহ, স্টেট-স্পেস মডেল (ম্যাম্বা), এনার্জি মিনিমাইজেশন ট্রান্সফরমার, জ্যামিতিক গ্রাফ কনভোলিউশন এবং স্কোর-ভিত্তিক জেনারেটিভ গতিবিদ্যা।

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: ব্যাকপ্রোপাগেশন, অপ্টিমাইজেশন ও সমতুল্যতা

খণ্ড ১ — লস ল্যান্ডস্কেপের জ্যামিতি — নন-কনভেক্স অপ্টিমাইজেশন ও স্যাডল পয়েন্ট অতিক্রম — 357 শব্দ
লস ল্যান্ডস্কেপের জ্যামিতি — নন-কনভেক্স অপ্টিমাইজেশন ও স্যাডল পয়েন্ট অতিক্রম

খণ্ড ২ — ব্যাকপ্রোপাগেশন ও অটোমেটিক ডিফারেন্সিয়েশন — কম্পিউটেশনাল গ্রাফ ও অ্যাডজয়েন্ট গ্রেডিয়েন্ট — 317 শব্দ
ব্যাকপ্রোপাগেশন ও অটোমেটিক ডিফারেন্সিয়েশন — কম্পিউটেশনাল গ্রাফ ও অ্যাডজয়েন্ট গ্রেডিয়েন্ট

খণ্ড ৩ — নরমালাইজেশন ও অপ্টিমাইজেশন গতিবিদ্যা — ব্যাচনর্ম, লেয়ারনর্ম, আরএমএসনর্ম এবং অ্যাডামডব্লিউ — 325 শব্দ
নরমালাইজেশন ও অপ্টিমাইজেশন গতিবিদ্যা — ব্যাচনর্ম, লেয়ারনর্ম, আরএমএসনর্ম এবং অ্যাডামডব্লিউ

খণ্ড ৪ — কনভোলিউশনাল সমতুল্যতা ও গ্রুপ থিওরি — লি প্রতিসাম্য ও ওজন অংশীদারিত্ব — 315 শব্দ
কনভোলিউশনাল সমতুল্যতা ও গ্রুপ থিওরি — লি প্রতিসাম্য ও ওজন অংশীদারিত্ব

দ্বিতীয় পর্যায়: রেসিডুয়াল গ্রেডিয়েন্ট ও অ্যাটেনশন মেকানিজম

খণ্ড ৫ — রেসিডুয়াল নেটওয়ার্ক ও অভেদ রূপান্তর — গ্রেডিয়েন্ট হাইওয়ে ও আনর‍্যাভেল্ড এনসেম্বল — 311 শব্দ
রেসিডুয়াল নেটওয়ার্ক ও অভেদ রূপান্তর — গ্রেডিয়েন্ট হাইওয়ে ও আনর‍্যাভেল্ড এনসেম্বল

খণ্ড ৬ — রিকারেন্ট সিকোয়েন্স ডায়নামিক্স — গ্রেডিয়েন্ট সংকট, এলএসটিএম সেল ও গেটেড মেমরি — 290 শব্দ
রিকারেন্ট সিকোয়েন্স ডায়নামিক্স — গ্রেডিয়েন্ট সংকট, এলএসটিএম সেল ও গেটেড মেমরি

খণ্ড ৭ — অ্যাটেনশন মেকানিজম ও কোয়েরি-কি-ভ্যালু দ্বৈততা — বাহদানাও থেকে স্কেল্ড ডট-প্রোডাক্ট — 276 শব্দ
অ্যাটেনশন মেকানিজম ও কোয়েরি-কি-ভ্যালু দ্বৈততা — বাহদানাও থেকে স্কেল্ড ডট-প্রোডাক্ট

খণ্ড ৮ — ট্রান্সফরমার আর্কিটেকচার — মাল্টি-হেড অ্যাটেনশন, পজিশনাল এনকোডিং ও রেসিডুয়াল স্ট্রিম — 267 শব্দ
ট্রান্সফরমার আর্কিটেকচার — মাল্টি-হেড অ্যাটেনশন, পজিশনাল এনকোডিং ও রেসিডুয়াল স্ট্রিম

তৃতীয় পর্যায়: স্টেট-স্পেস, গ্রাফ ম্যানিফোল্ড ও এনার্জি মডেল

খণ্ড ৯ — রোটারি পজিশন এম্বেডিং (RoPE) ও কনটেক্সট উইন্ডো — জটিল ঘূর্ণন ও দীর্ঘ দূরত্বের বিস্তার — 319 শব্দ
রোটারি পজিশন এম্বেডিং (RoPE) ও কনটেক্সট উইন্ডো — জটিল ঘূর্ণন ও দীর্ঘ দূরত্বের বিস্তার

খণ্ড ১০ — অটোএনকোডার ও ল্যাটেন্ট রিপ্রেজেন্টেশন — ম্যানিফোল্ড বটলনেক ও ডিসএন্ট্যাঙ্গলমেন্ট — 258 শব্দ
অটোএনকোডার ও ল্যাটেন্ট রিপ্রেজেন্টেশন — ম্যানিফোল্ড বটলনেক ও ডিসএন্ট্যাঙ্গলমেন্ট

খণ্ড ১১ — ভ্যারিয়েশনাল অটোএনকোডার (VAE) — এভিডেন্স লোয়ার বাউন্ড ও রিপ্যারামিটারাইজেশন ট্রিক — 263 শব্দ
ভ্যারিয়েশনাল অটোএনকোডার (VAE) — এভিডেন্স লোয়ার বাউন্ড ও রিপ্যারামিটারাইজেশন ট্রিক

খণ্ড ১২ — জেনারেটিভ অ্যাডভারসারিয়াল নেটওয়ার্ক (GAN) — মিনিম্যাক্স গেম ও ওয়াসারস্টাইন জ্যামিতি — 285 শব্দ
জেনারেটিভ অ্যাডভারসারিয়াল নেটওয়ার্ক (GAN) — মিনিম্যাক্স গেম ও ওয়াসারস্টাইন জ্যামিতি

চতুর্থ পর্যায়: ভ্যারিয়েশনাল ও স্কোর-ভিত্তিক জেনারেটিভ মডেল

খণ্ড ১৩ — নরমালাইজিং ফ্লো ও উভমুখী নিউরাল নেটওয়ার্ক — চলক রূপান্তর ও নিখুঁত সম্ভাবনা — 287 শব্দ
নরমালাইজিং ফ্লো ও উভমুখী নিউরাল নেটওয়ার্ক — চলক রূপান্তর ও নিখুঁত সম্ভাবনা

খণ্ড ১৪ — এনার্জি-ভিত্তিক মডেল ও আধুনিক হপফিল্ড নেটওয়ার্ক — বোল্টজম্যান মেশিন থেকে ডেনস মেমরি — 310 শব্দ
এনার্জি-ভিত্তিক মডেল ও আধুনিক হপফিল্ড নেটওয়ার্ক — বোল্টজম্যান মেশিন থেকে ডেনস মেমরি

খণ্ড ১৫ — গ্রাফ নিউরাল নেটওয়ার্ক (GNN) — মেসেজ পাসিং ও ওয়েইসফেইলার-লেহম্যানের সীমা — 285 শব্দ

Shakil Mahmud • Research Monograph

গ্রাফ নিউরাল নেটওয়ার্ক (GNN) — মেসেজ পাসিং ও ওয়েইসফেইলার-লেহম্যানের সীমা

খণ্ড ১৬ **জ্যামিতিক ডিপ লার্নিং ও স্ফেরিক্যাল সিএনএন — গেজ সিমেট্রি ও ত্রিমাত্রিক ঘূর্ণন সমতুল্যতা** 285 শব্দ
জ্যামিতিক ডিপ লার্নিং ও স্ফেরিক্যাল সিএনএন — গেজ সিমেট্রি ও ত্রিমাত্রিক ঘূর্ণন সমতুল্যতা

খণ্ড ১৭ **নিউরাল অর্ডিনারি ডিফারেনশিয়াল সমীকরণ — অবিচ্ছিন্ন গভীরতা ও অ্যাডজয়েন্ট পদ্ধতি** 310 শব্দ
নিউরাল অর্ডিনারি ডিফারেনশিয়াল সমীকরণ — অবিচ্ছিন্ন গভীরতা ও অ্যাডজয়েন্ট পদ্ধতি

খণ্ড ১৮ **স্টেট-স্পেস মডেল ও ম্যাম্বা — হিপো প্রজেকশন ও লিনিয়ার-টাইম সিকোয়েন্স** 310 শব্দ
স্টেট-স্পেস মডেল ও ম্যাম্বা — হিপো প্রজেকশন ও লিনিয়ার-টাইম সিকোয়েন্স

খণ্ড ১৯ **ডিপ রিইনফোর্সমেন্ট লার্নিং আর্কিটেকচার — ডিপ কিউ-নেটওয়ার্ক থেকে অ্যাক্টর-ক্রিটিক দ্বৈততা** 304 শব্দ
ডিপ রিইনফোর্সমেন্ট লার্নিং আর্কিটেকচার — ডিপ কিউ-নেটওয়ার্ক থেকে অ্যাক্টর-ক্রিটিক দ্বৈততা

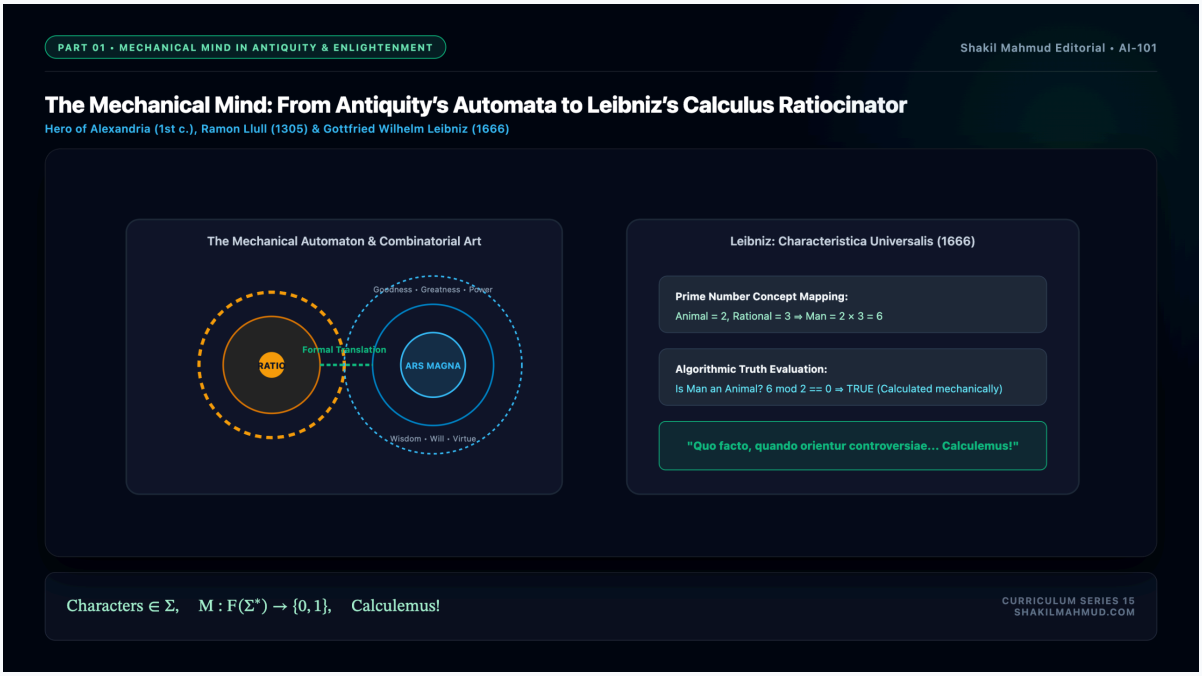
খণ্ড ২০ **ডিপ নেটওয়ার্কের রিপ্রেজেন্টেশন তত্ত্ব — ইনফরমেশন বটলনেক ও নিউরাল কোলাপ্স** 341 শব্দ
ডিপ নেটওয়ার্কের রিপ্রেজেন্টেশন তত্ত্ব — ইনফরমেশন বটলনেক ও নিউরাল কোলাপ্স

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • প্রথম খণ্ড

লস ল্যান্ডস্কেপের জ্যামিতি — নন-কনভেক্স অপ্টিমাইজেশন ও স্যাডল পয়েন্ট অতিক্রম

লস ল্যান্ডস্কেপের জ্যামিতি — নন-কনভেক্স অপ্টিমাইজেশন ও স্যাডল পয়েন্ট অতিক্রম

357 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।*

চিরায়ত অপ্টিমাইজেশন তত্ত্বের মূল ভিত্তি ছিল উত্তল (Convex) ফাংশন, যেখানে যেকোনো স্থানীয় সর্বনিম্ন বিন্দু (Local Minimum) স্বতঃসিদ্ধভাবে বৈশ্বিক সর্বনিম্ন বিন্দু (Global Minimum)। কিন্তু গভীর নিউরাল নেটওয়ার্কের জটিল প্যারামিটার স্পেস $mathbbR^D$-তে যেখানে মাত্রা D sim 10^7 – 10^{11}, সেখানে লস ল্যান্ডস্কেপ চরম নন-কনভেক্স। একসময় ধারণা করা হতো গ্রেডিয়েন্ট ডিসেন্ট স্থানীয় ফাঁদে আটকা পড়ে যাবে। ২০১৪ সালে ইয়ান ডফাঁ এবং সহকর্মীরা স্পিন-গ্লাসের পরিসংখ্যানিক বলবিদ্যা প্রয়োগ করে প্রমাণ করেন: উচ্চ মাত্রায় উচ্চ লসের অঞ্চলে লোকাল মিনিমা প্রায় নেই বললেই চলে; বরং সেখানে আধিপত্য বিস্তার করে থাকে অগণিত **হাইপারবোলিক স্যাডল পয়েন্ট (Saddle Points)**।

১. উচ্চ-মাত্রিক ক্রিটিক্যাল পয়েন্টের টপোলজি

প্যারামিটার ভেক্টর $thetainmathbbR^D$-এর জন্য ক্রিটিক্যাল পয়েন্টে প্রথম অন্তরক শূন্য হয়: $nabla_thetaL(theta) = 0$। এই বিন্দুর স্থানীয় জ্যামিতি নির্ধারিত হয় দ্বিতীয় ঘাতের **হেসিয়ান ম্যাট্রিক্স (Hessian Matrix)** $H_ij = fracpartial^2 Lpartialtheta_ipartialtheta_j$-এর মাধ্যমে:

$$L(theta + Deltatheta)approxL(theta) + frac12Deltatheta^T H(theta)Deltatheta$$

মোর্স ইনডেক্স (Morse Index) $alpha$ হলো হেসিয়ানের ঋণাত্মক আইগেনভ্যালুসমূহের অনুপাত। একটি ক্রিটিক্যাল পয়েন্টকে লোকাল মিনিমাম হতে হলে তার সমস্ত D সংখ্যক আইগেনভ্যালুকেই নিশ্চিতভাবে ধনাত্মক হতে হবে। যদি আইগেনভ্যালুগুলোর চিহ্ন স্বাধীন মুদ্রা টসের মতো আচরণ করে, তবে কোনো পয়েন্টের লোকাল মিনিমাম হওয়ার সম্ভাবনা দাঁড়ায়:

Shakil Mahmud • Research Monograph

$$mathbbP(textlocalminimum) = 2^{-D}$$

লক্ষাধিক মাত্রার নেটওয়ার্কে এই সম্ভাবনা কার্যত শূন্যের কোঠায় নেমে আসে, যা প্রমাণ করে যে উচ্চ লস স্তরে লোকাল মিনিমার কোনো বাস্তবিক অস্তিত্ব নেই।

২. র‍্যান্ডম ম্যাট্রিক্স তত্ত্ব ও হেসিয়ান স্পেকট্রাম

আধুনিক ডিপ লার্নিংয়ে **র‍্যান্ডম ম্যাট্রিক্স থিওরি (RMT)** প্রয়োগ করে দেখা যায় যে হেসিয়ানের স্পেকট্রাম দুটি ভাগে বিভক্ত:

- **কন্টিনিউয়াস বাল্ক (Continuous Bulk):** মারচেঙ্কো-পাস্তুর বা উইগনার অর্ধবৃত্ত নিয়ম মেনে চলা প্রায়-শূন্য আইগেনভ্যালুসমূহের বিশাল পুঞ্জ। এগুলো ল্যান্ডস্কেপের সমতল ও নিষ্ক্রিয় দিকগুলোকে নির্দেশ করে।
- **বিচ্ছিন্ন শীর্ষ স্পাইক (Isolated Spikes):** অল্প কয়েকটি বিশাল ধনাত্মক আইগেনভ্যালু ($lambda_{max} gg 0$) যা মডেলের মূল টাস্ক বা রিপ্রেজেন্টেশনের অভিমুখ নির্দেশ করে।

৩. স্যাডল পয়েন্ট থেকে স্টোক্যাস্টিক উত্তরণ

যদি কোনো স্যাডল পয়েন্টে অন্তত একটি ঋণাত্মক আইগেনভ্যালু থাকে (lambda_{min} < 0), তবে তাকে *স্ট্রিক্ট স্যাডল পয়েন্ট* বলে। ২০১৫ সালে রং গে এবং ২০১৭ সালে চি জিন প্রমাণ করেন যে **পার্টার্বড স্টোক্যাস্টিক গ্রেডিয়েন্ট ডিসেন্ট (Perturbed SGD)** যেকোনো স্ট্রিক্ট স্যাডল পয়েন্ট থেকে বহুপদী সময়ে (Polynomial Time) সফলভাবে মুক্ত হতে পারে। স্টোক্যাস্টিক নয়েজ ঋণাত্মক বক্রতার দিকে ক্ষুদ্র ধাক্কা দেয়, যার ফলে গতিশীল সিস্টেমটি অসমতা ভেঙে দ্রুত নিম্নতর উপত্যকায় পতিত হয়।

সাইটেশন: মাহমুদ, শাকিল (2026)। "লস ল্যান্ডস্কেপের জ্যামিতি — নন-কনভেক্স অপ্টিমাইজেশন ও স্যাডল পয়েন্ট অতিক্রম"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১।
https://shakilmahmud.com/bn/deep-learning-architectures-part-01-loss-landscape-geometry-saddle-points-bn/

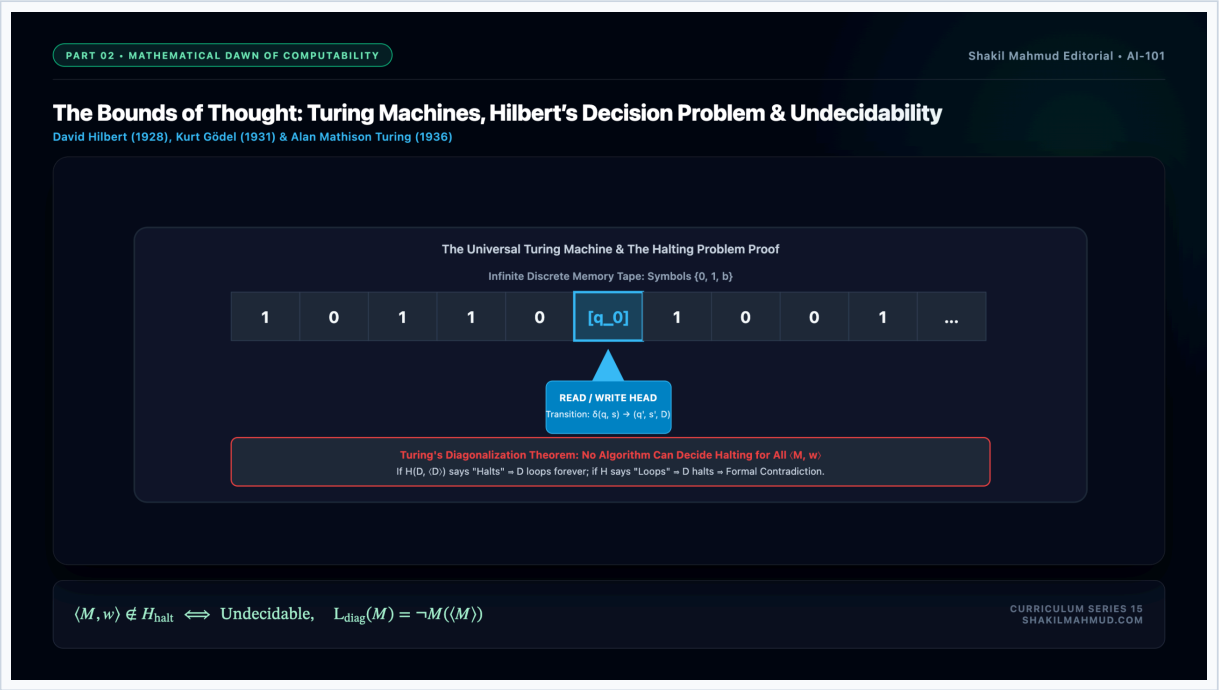
শাকিল মাহমুদ মনোগ্রাফ সিরিজ • দ্বিতীয় খণ্ড

ব্যাকপ্রোপাগেশন ও অটোমেটিক ডিফারেন্সিয়েশন — কম্পিউটেশনাল গ্রাফ ও অ্যাডজয়েন্ট গ্রেডিয়েন্ট

ব্যাকপ্রোপাগেশন ও অটোমেটিক ডিফারেন্সিয়েশন — কম্পিউটেশনাল গ্রাফ ও অ্যাডজয়েন্ট গ্রেডিয়েন্ট

317 শব্দ • November 4, 2026 • ওয়েব সংস্করণ

❖



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

আধুনিক কৃত্রিম বুদ্ধিমত্তার ইঞ্জিনের প্রধানতম গাণিতিক ভিত্তি হলো **অটোমেটিক ডিফারেন্সিয়েশন (Automatic Differentiation – AD)**—যার রিভার্স-মোড সংস্করণটি মেশিন লার্নিং জগতে বিখ্যাত **ব্যাকপ্রোপাগেশন (Backpropagation)** অ্যালগরিদম হিসেবে পরিচিত। নিউমেরিক্যাল ডিফারেন্সিয়েশনের মতো এর কোনো রাউন্ড-অফ ত্রুটি বা বিশাল কম্পিউটেশনাল জটিলতা ($O(D)$) নেই, আবার সিম্বলিক ডিফারেন্সিয়েশনের মতো সমীকরণ অতিরিক্ত দীর্ঘায়িত হওয়ার ঝুঁকিও নেই। ব্যাকপ্রোপাগেশন একটি মাত্র ফরোয়ার্ড ও ব্যাকওয়ার্ড পাসের মাধ্যমে কোটি কোটি প্যারামিটারের নিখুঁত গ্রেডিয়েন্ট নির্ণয় করে $O(1)$ আনুপাতিক সময়ে।

১. কম্পিউটেশনাল ডিরেক্টেড অ্যাসাইক্লিক গ্রাফ (DAG)

যেকোনো জটিল গাণিতিক অপারেশনকে কতগুলো প্রাথমিক অপারেশনের সমন্বয়ে একটি নির্দেশিত চক্রবিহীন গ্রাফে (DAG) রূপান্তর করা যায়। গ্রাফের নোডগুলো টপোলজিক্যাল ক্রমে সাজানো থাকে: ইনপুট নোডগুলো থেকে শুরু করে ক্রমান্বয়ে মধ্যবর্তী নোড হয়ে চূড়ান্ত নোডে স্কেলার লস L উৎপন্ন হয়।

২. ফরোয়ার্ড বনাম রিভার্স মোডের মাত্রাগত দ্বৈততা

একটি সাধারণ রূপান্তর $f : mathbbR^n tomathbbR^m$-এর জন্য প্রথম-ঘাতের সমস্ত আংশিক অন্তরক ধারণ করে জ্যাকোবিয়ান ম্যাট্রিক্স $JinmathbbR^{mtimesn}$।

- **ফরোয়ার্ড-মোড এডি:** জ্যাকোবিয়ান-ভেক্টর প্রোডাক্ট ($Jcdotv$) নির্ণয় করে। এটি উপযোগী যখন ইনপুট ডাইমেনশন আউটপুট অপেক্ষা অনেক কম ($nllm$)।

Shakil Mahmud • Research Monograph

- **রিভার্স-মোড এডি:** ভেক্টর-জ্যাকোবিয়ান প্রোডাক্ট ($u^T cdot J$) নির্ণয় করে। ডিপ লার্নিংয়ে স্কেলার লস হওয়ায় আউটপুট মাত্রা $m = 1$ কিন্তু প্যারামিটার সংখ্যা n = D sim 10^8 – 10^{11}। ফলে রিভার্স-মোড মাত্র একটি মাত্র পদক্ষেপে সমগ্র প্যারামিটার স্পেসের গ্রেডিয়েন্ট $nabla_{t}hetaL$ নিখুঁতভাবে বের করে দেয়।

৩. ব্যাকপ্রোপাগেশনের অ্যাডজয়েন্ট সমীকরণ

যেকোনো নোড v_i-এর অ্যাডজয়েন্ট মানকে সংজ্ঞায়িত করা হয় চূড়ান্ত লস L-এর সাপেক্ষে তার সংবেদনশীলতা হিসেবে: $barv_{i}equivfracpartialLpartialv_{i}$। মাল্টিভ্যারিয়েট চেইন রুল অনুসারে:

$$barv_{i} = sum_{jintextChildren(v_{i})}barv_{j}fracpartialphi_{j}partialv_{i}$$

টার্মিনাল নোডে $barL = 1.0$ ধরে গ্রাফের বিপরীতমুখী টপোলজিক্যাল যাত্রার মাধ্যমে প্রতিটি প্যারামিটার নোডে অ্যাডজয়েন্ট গ্রেডিয়েন্ট সংগৃহীত হয়।

৪. মেমরি সেভিং: গ্রেডিয়েন্ট চেকপয়েন্টিং

ব্যাকওয়ার্ড পাসে ব্যবহারের জন্য ফরোয়ার্ড পাসের সমস্ত ইন্টারমিডিয়েট অ্যাক্টিভেশন মেমরিতে জমিয়ে রাখতে হয়, যার ফলে মেমরি খরচ দাঁড়ায় $O(N)$। ২০১৬ সালে চেন উদ্ভাবন করেন **গ্রেডিয়েন্ট চেকপয়েন্টিং**: কেবল নির্দিষ্ট দূরত্বের চেকপয়েন্ট নোডগুলো সংরক্ষণ করে ব্যাকওয়ার্ড পাসের সময় তাৎক্ষণিকভাবে মধ্যবর্তী মানগুলো পুনরায় গণনা করার মাধ্যমে মেমরি খরচ $O(N)$ থেকে নাটকীয়ভাবে $O(sqrtN)$-এ নামিয়ে আনা হয়।

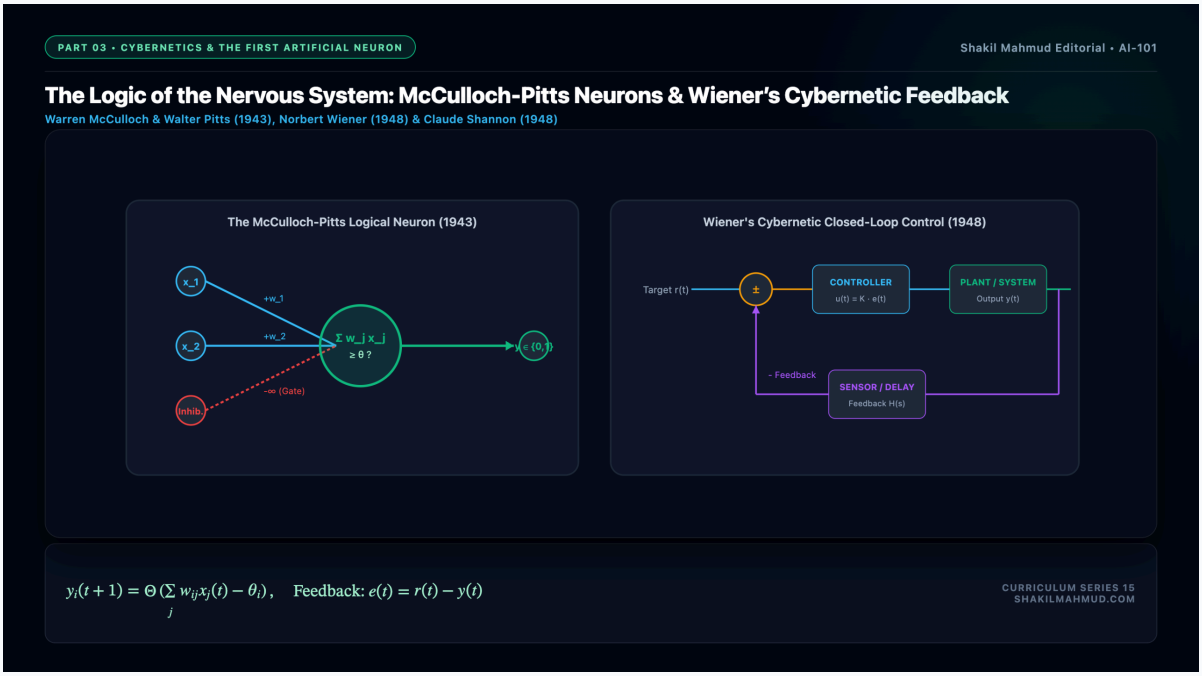
সাইটেশন: মাহমুদ, শাকিল (2026)। "ব্যাকপ্রোপাগেশন ও অটোমেটিক ডিফারেন্সিয়েশন — কম্পিউটেশনাল গ্রাফ ও অ্যাডজয়েন্ট গ্রেডিয়েন্ট"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ২।
https://shakilmahmud.com/bn/deep-learning-architectures-part-02-backpropagation-automatic-differentiation-computational-graphs-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • তৃতীয় খণ্ড

নরমালাইজেশন ও অপ্টিমাইজেশন গতিবিদ্যা — ব্যাচনর্ম, লেয়ারনর্ম, আরএমএসনর্ম এবং অ্যাডামডব্লিউ

নরমালাইজেশন ও অপ্টিমাইজেশন গতিবিদ্যা — ব্যাচনর্ম, লেয়ারনর্ম, আরএমএসনর্ম এবং অ্যাডামডব্লিউ

325 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



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

ডিপ নিউরাল নেটওয়ার্কে শত শত স্তর যুক্ত করার প্রচেষ্টা একসময় মারাত্মক সংখ্যাগত অস্থিতিশীলতার শিকার হতো: এক্সপ্লোডিং বা ভ্যানিশিং গ্রেডিয়েন্ট, অস্বাভাবিক ভ্যারিয়েন্স এবং সূচনার প্রতি চরম সংবেদনশীলতা। আধুনিক ডিপ লার্নিং বিপ্লব সম্ভব হয়েছে মূলত কার্যকর **নরমালাইজেশন স্তর** এবং বক্রতা-সংবেদনশীল অপ্টিমাইজার **অ্যাডামডব্লিউ (AdamW)**-এর যৌথ অবদানে। এই কৌশলগুলো বন্ধুর, এবড়োখেবড়ো অপ্টিমাইজেশন ল্যান্ডস্কেপকে একটি মসৃণ ও সুষম জ্যামিতিক তলে রূপান্তরিত করে।

১. ব্যাচ নরমালাইজেশন: ল্যান্ডস্কেপ স্মুথিং

২০১৫ সালে সের্গেই আইওফে এবং ক্রিশ্চিয়ান সেগেডি **ব্যাচ নরমালাইজেশন (BatchNorm)** প্রবর্তন করেন। মিনি-ব্যাচ $mathcal B$-এর জন্য এটি প্রতিটি চ্যানেলকে স্বাধীনভাবে গড় ও ভ্যারিয়েন্স দিয়ে মানসম্মত করে affine প্যারামিটার $gamma$ ও $beta$ প্রয়োগ করে:

hat{x}_i = frac{x_i – mu_{mathcal{B}}}{sqrt{sigma_{mathcal{B}}^2 + epsilon}}, quad y_i = gamma hat{x}_i + beta

২০১৮ সালে সান্তুরকার প্রমাণ করেন যে ব্যাচনর্মের মূল সাফল্য কোভ্যারিয়েট শিফট দূর করায় নয়, বরং ল্যান্ডস্কেপের লিপশিটজ কনস্ট্যান্ট নিয়ন্ত্রণ করে হেসিয়ানের কন্ডিশন নাম্বার সংকুচিত করার মাঝে নিহিত।

Shakil Mahmud • Research Monograph

২. লেয়ারনর্ম এবং আরএমএসনর্ম: সিকোয়েন্স মডেলের ভিত্তি

কম্পিউটার ভিশনে ব্যাচনর্ম সেরা হলেও অটোরিগ্রেসিভ সিকোয়েন্স মডেলে এটি ব্যর্থ হয়। এর সমাধানে আসে:

- **লেয়ার নরমালাইজেশন (LayerNorm):** প্রতিটি নমুনার জন্য তার সমস্ত অভ্যন্তরীণ ফিচার বা চ্যানেল ডাইমেনশন বরাবর স্বাধীনভাবে গড় ও ভ্যারিয়েন্স হিসাব করে।
- **আরএমএসনর্ম (RMSNorm):** ২০১৯ সালে ঝাং দেখান যে গড় বিয়োগ করা অপ্রয়োজনীয়; কেবল রুট মিন স্কয়ার এনার্জি $textRMS(x) = sqrtfrac1Dsumx_j^2 + epsilon$ দিয়ে ভাগ করাই যথেষ্ট। এটি আধুনিক এলএলএম (LLaMA, Mistral, Gemma)-এ সময় ও মেমরি সাশ্রয়ী রূপ হিসেবে ব্যবহৃত হচ্ছে।

৩. প্রি-এলএন (Pre-LN) বনাম পোস্ট-এলএন (Post-LN)

মূল ট্রান্সফরমারের পোস্ট-এলএন ব্যবস্থায় রেসিডুয়াল কানেকশনের বাইরে নরমালাইজেশন করায় প্রাথমিক স্তরগুলোতে গ্রেডিয়েন্ট সংকুচিত হয়ে যেত, যার ফলে কঠোর ওয়ার্মআপের প্রয়োজন হতো। আধুনিক প্রি-এলএন আর্কিটেকচারে নরমালাইজেশনকে সাব-লেয়ারের ভেতরে স্থাপন করায় মূল রেসিডুয়াল স্ট্রিম সম্পূর্ণ উন্মুক্ত থাকে, ফলে শত শত স্তর গভীর হলেও গ্রেডিয়েন্ট অবিকল প্রবাহিত হয়।

৪. অ্যাডামডব্লিউ (AdamW): ডিকাপল্ড ওয়েট ডিকে

ক্লাসিক্যাল অ্যাডামে L2 রেগুলাইজেশন সরাসরি লসের সাথে যুক্ত করায় অ্যাডাপ্টিভ প্রিকন্ডিশনার $sqrtv_t$ দ্বারা ওয়েট ডিকে ভাগ হয়ে বিকৃত হয়ে যেত। ২০১৭ সালে লশচিলভ ও হুটার **অ্যাডামডব্লিউ** উদ্ভাবন করেন, যা গ্রেডিয়েন্ট স্টেপ নেওয়ার পর সরাসরি ইউক্লিডীয় সংকোচন ঘটায় (theta_{t+1} = theta_t – eta lambda theta_t), যা মডেলের সাধারণীকরণ ক্ষমতা বহুগুণ বাড়িয়ে দেয়।

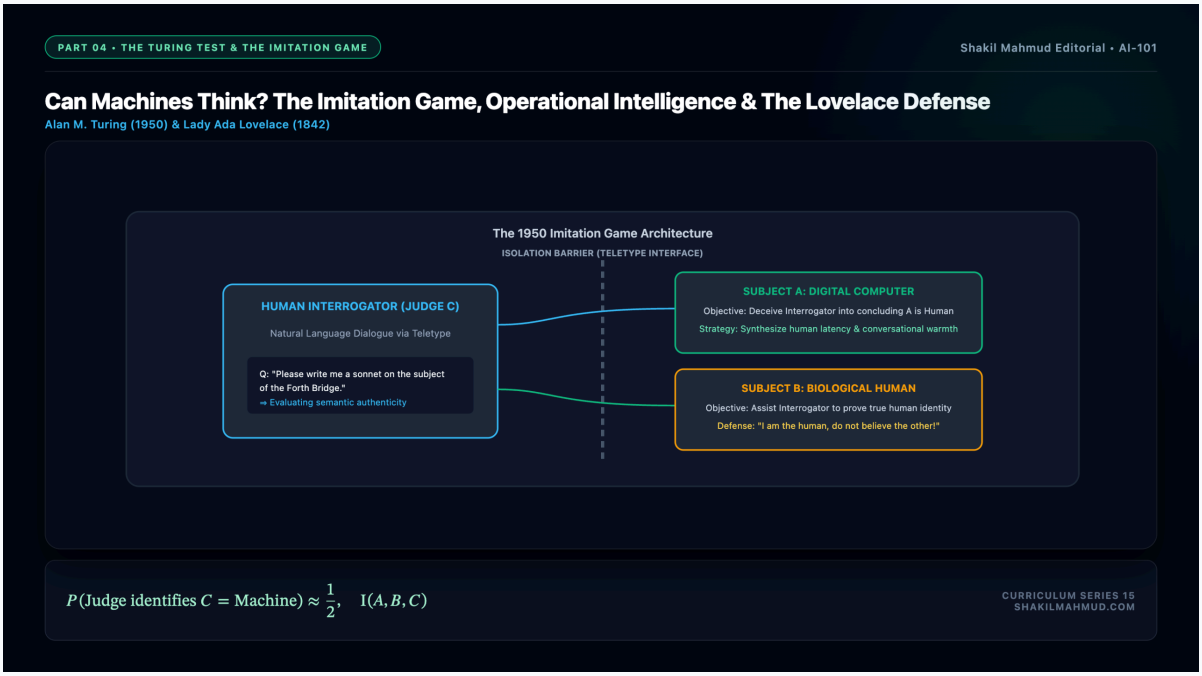
সাইটেশন: মাহমুদ, শাকিল (2026)। "নরমালাইজেশন ও অপ্টিমাইজেশন গতিবিদ্যা — ব্যাচনর্ম, লেয়ারনর্ম, আরএমএসনর্ম এবং অ্যাডামডব্লিউ"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ৩। https://shakilmahmud.com/bn/deep-learning-architectures-part-03-normalization-batchnorm-layernorm-rmsnorm-adamw-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • চতুর্থ খণ্ড

কনভোলিউশনাল সমতুল্যতা ও গ্রুপ থিওরি — লি প্রতিসাম্য ও ওজন অংশীদারিত্ব

কনভোলিউশনাল সমতুল্যতা ও গ্রুপ থিওরি — লি প্রতিসাম্য ও ওজন অংশীদারিত্ব

315 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।*

১৮৭২ সালে ফেলিক্স ক্লাইন তার বিখ্যাত *এরলাঙ্গেন প্রোগ্রাম (Erlangen Program)*-এ প্রস্তাব করেছিলেন যে জ্যামিতি কেবল দূরত্বের পাঠ নয়, বরং বিভিন্ন রূপান্তর গ্রুপের অধীনে অপরিবর্তনীয় বৈশিষ্ট্যসমূহের অনুসন্ধান। আধুনিক রিপ্রেজেন্টেশন তত্ত্বে এই অন্তর্দৃষ্টিই হলো **কনভোলিউশনাল নিউরাল নেটওয়ার্ক (CNN)**-এর গাণিতিক ভিত্তি। ত্রিমাত্রিক বা দ্বিমাত্রিক স্থানের স্থানান্তর প্রতিসাম্য গ্রুপকে $(mathbbR^2, +)$ কাজে লাগিয়ে কনভোলিউশন **সমতুল্যতা বা ইকুইভ্যারিয়েন্স (Equivariance)** প্রতিষ্ঠা করে, যা প্যারামিটার জটিলতাকে সম্পূর্ণ সংযুক্ত স্তরের $O(D^2)$ থেকে ফিল্টার আকারের $O(K^2)$-এ সংকুচিত করে।

১. গ্রুপ ইকুইভ্যারিয়েন্সের আনুষ্ঠানিক সংজ্ঞা

ধরি একটি রূপান্তর গ্রুপ G ইনপুট স্পেসে T_g এবং আউটপুট স্পেসে T_g’ হিসেবে ক্রিয়া করে। একটি ফাংশন f-কে G-**Equivariant** বলা হবে যদি:

$$f(T_g x) = T_g’ (f(x)), quad forall x in mathcal{X}, ; forall g in G$$

অর্থাৎ ইনপুট স্থানান্তরিত হলে আউটপুটও সমভাবে স্থানান্তরিত হবে। গভীর দৃষ্টি নেটওয়ার্কে প্রাথমিক স্তরগুলো স্থানিক সমতুল্যতা (Equivariance) বজায় রাখে যাতে বস্তুর সুনির্দিষ্ট অবস্থান শনাক্ত করা যায়, আর শীর্ষ স্তরগুলো ইনভ্যারিয়েন্স (Invariance) অর্জন করে যাতে অবস্থান নির্বিশেষে বস্তুর পরিচয় নিশ্চিত করা যায়।

২. স্থানান্তর সমতুল্যতা ও কনভোলিউশন উপপাদ্য

অবিচ্ছিন্ন স্থানান্তর অপারেটর T_t f(x) = f(x – t)-এর জন্য কনভোলিউশন ইন্টিগ্রাল বিবেচনা করি:

Shakil Mahmud • Research Monograph

(f * w)(x) = int_{mathbb{R}^d} f(y) w(x – y) dy

চলক প্রতিস্থাপনের মাধ্যমে দেখানো যায় যে $(T_t f * w)(x) = T_t(f * w)(x)$। লিনিয়ার সিস্টেম তত্ত্বের মৌলিক উপপাদ্য অনুযায়ী: *যেকোনো লিনিয়ার অপারেটর যা স্থানান্তরের সাথে সমতুল্য (Equivariant), তা অবশ্যই একটি কনভোলিউশন অপারেটর হতে বাধ্য*। সুতরাং ওজন অংশীদারিত্ব (Weight Sharing) কোনো আন্দাজি কৌশল নয়, এটি স্থানান্তর প্রতিসাম্যের অনিবার্য গাণিতিক পরিণতি।

৩. গ্রুপ কনভোলিউশনাল নেটওয়ার্ক (G-CNN)

২০১৬ সালে টাকো কোহেন এবং ম্যাক্স ওয়েলিং কনভোলিউশনকে ঘূর্ণন ও প্রতিফলন গ্রুপে ($SE(2)$) প্রসারিত করেন। হার মেজার (Haar Measure) dg -এর মাধ্যমে গ্রুপ কনভোলিউশন সংজ্ঞায়িত করে G-CNN আবর্তনের বিরুদ্ধেও নিখুঁত সমতুল্যতা অর্জন করে।

৪. ফুরিয়ার রূপান্তর ও দ্বৈততা

কনভোলিউশন উপপাদ্য প্রমাণ করে যে স্থানিক ডোমেনের কনভোলিউশন হলো ফ্রিকোয়েন্সি ডোমেনের গুণফল: $mathcal{F} f * w = hat{f} cdot hat{w}$। এই গাণিতিক দ্বৈততা প্রকাশ করে যে প্রতিটি কনভোলিউশন লেয়ার আসলে একটি ফিল্টার ব্যাংক হিসেবে নির্দিষ্ট তরঙ্গ ফ্রিকোয়েন্সিকে পৃথক ও প্রক্রিয়াজাত করে।

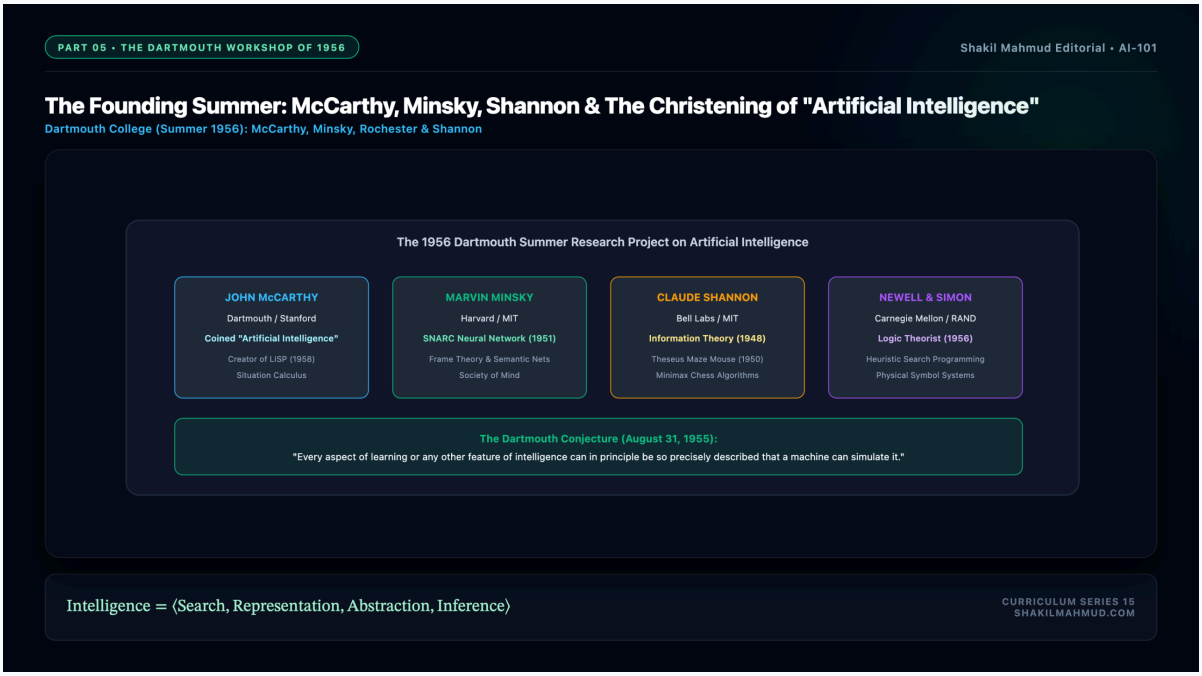
সাইটেশন: মাহমুদ, শাকিল (2026)। "কনভোলিউশনাল সমতুল্যতা ও গ্রুপ থিওরি — লি প্রতিসাম্য ও ওজন অংশীদারিত্ব"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ৪।
https://shakilmahmud.com/bn/deep-learning-architectures-part-04-convolutional-equivariance-group-theory-lie-symmetries-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • পঞ্চম খণ্ড

রেসিডুয়াল নেটওয়ার্ক ও অভেদ রূপান্তর — গ্রেডিয়েন্ট হাইওয়ে ও আনর‍্যাভেল্ড এনসেম্বল

রেসিডুয়াল নেটওয়ার্ক ও অভেদ রূপান্তর — গ্রেডিয়েন্ট হাইওয়ে ও আনর‍্যাভেল্ড এনসেম্বল

311 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



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

২০১৫ সালের পূর্বে ডিপ লার্নিং গবেষকরা একটি অপ্রত্যাশিত বাধার সম্মুখীন হন যাকে বলা হতো **ডিগ্রোডেশন সমস্যা (Degradation Problem)**। নেটওয়ার্কের গভীরতা ২০ স্তর অতিক্রম করলেই ট্রেনিং এরর—কেবল টেস্ট এরর নয়—হঠাৎ অস্বাভাবিক হারে বৃদ্ধি পেত। এটি কোনো ওভারফিটিং ছিল না (ওভারফিটিংয়ে ট্রেনিং এরর কমে কিন্তু টেস্ট এরর বাড়ে); এটি ছিল গভীর স্তরে শেখার চরম ব্যর্থতা। এমনকি সাধারণ অভেদ ফাংশন ($mathcal{H}(x) = x$) শেখাও গভীর স্তরের জন্য অসম্ভব হয়ে পড়ত। ২০১৫ সালের ডিসেম্বরে কাইমিং হে এবং তার দল **রেসিডুয়াল নেটওয়ার্ক (ResNet)** আবিষ্কারের মাধ্যমে এই সংকট দূর করেন এবং ১০০০ স্তরেরও বেশি গভীর মডেল সফলভাবে প্রশিক্ষণের পথ উন্মোচন করেন।

১. রেসিডুয়াল কাঠামোর ধারণা

হে এবং তার সহকর্মীরা বুঝতে পেরেছিলেন যে একটি স্তরকে সরাসরি জটিল ম্যাপিং $mathcal{H}(x)$ শেখার নির্দেশ না দিয়ে, তাকে কেবল অবশিষ্ট বা **রেসিডুয়াল ম্যাপিং** mathcal{F}(x) equiv mathcal{H}(x) – x শেখানো উচিত:

$$mathcal{H}(x) = mathcal{F}(x) + x$$

যদি কোনো স্তরে নতুন কোনো তথ্য যোগ করার প্রয়োজন না থাকে, তবে রেগুলেশনের মাধ্যমে সহজেই তার প্যারামিটারগুলো শূন্যের দিকে ধাবিত হবে ($mathcal{F}(x) to 0$) এবং ইনপুট সরাসরি আউটপুটে অপরিবর্তিত অবস্থায় পৌঁছে যাবে।

Shakil Mahmud • Research Monograph

২. বাধাহীন গ্রেডিয়েন্ট হাইওয়ের গাণিতিক উপপাদন

রেসিডুয়াল ব্লকের সমীকরণ $x_{l+1} = x_l + mathcal{F}(x_l, mathcal{W}_l)$-কে পুনরাবৃত্তি করলে যেকোনো গভীর স্তর L-এর মান দাঁড়ায়:

$$x_L = x_l + sum_{i=l}^{L-1} mathcal{F}(x_i, mathcal{W}_i)$$

চেইন রুল প্রয়োগের মাধ্যমে লস $mathcal{E}$-এর গ্রেডিয়েন্ট নির্ণয় করলে দেখা যায়:

$$frac{partial mathcal{E}}{partial x_l} = frac{partial mathcal{E}}{partial x_L} left(I + frac{partial}{partial x_l} sum_{i=l}^{L-1} mathcal{F}_i right)$$

এই সমীকরণে একক ম্যাট্রিক্স I একটি **গ্রেডিয়েন্ট হাইওয়ে** হিসেবে কাজ করে। এমনকি মধ্যবর্তী নন-লিনিয়ার লেয়ারগুলোর গ্রেডিয়েন্ট শূন্য হয়ে গেলেও, এই অভেদ পদের কারণে শীর্ষ স্তরের সিগন্যাল কোনো ক্ষয় ছাড়াই সরাসরি প্রাথমিক স্তরগুলোতে অবিকল পৌঁছে যায়।

৩. আনর‍্যাভেল্ড এনসেম্বল তত্ত্ব (Unraveled Ensemble)

২০১৬ সালে ভাইট ও তার সহকর্মীরা প্রমাণ করেন যে রেসনেট আসলে একটি একক অতি-গভীর নেটওয়ার্ক নয়, বরং এটি 2^n সংখ্যক অগভীর নেটওয়ার্কের সমান্তরাল সমন্বয়। যেকোনো একটি স্তর মুছে ফেললেও সামগ্রিক কার্যকারিতায় কোনো বিপর্যয় ঘটে না, যা প্রমাণ করে যে এটি একটি স্বতঃস্ফূর্ত এনসেম্বল সিস্টেম হিসেবে কাজ করে।



সাইটেশন: মাহমুদ, শাকিল (2026)। "রেসিডুয়াল নেটওয়ার্ক ও অভেদ রূপান্তর — গ্রেডিয়েন্ট হাইওয়ে ও আনর‍্যাভেল্ড এনসেম্বল"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ৫। https://shakilmahmud.com/bn/deep-learning-architectures-part-05-residual-networks-identity-mappings-gradient-highways-bn/

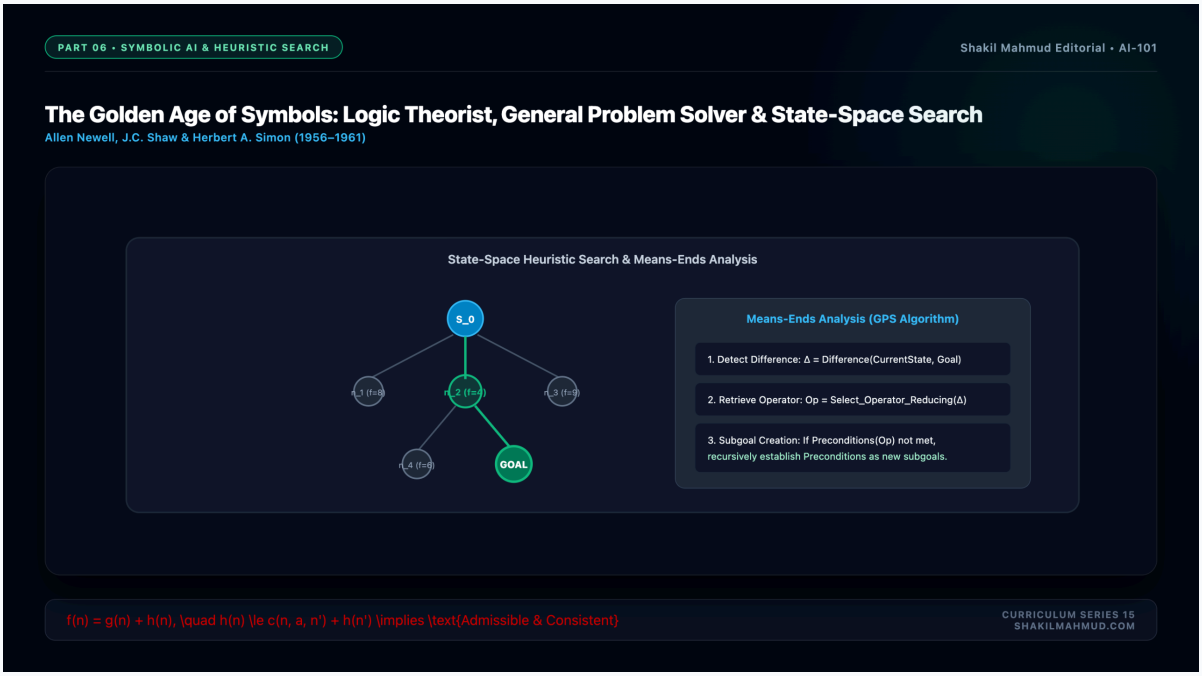
শাকিল মাহমুদ মনোগ্রাফ সিরিজ • ষষ্ঠ খণ্ড

রিকারেন্ট সিকোয়েন্স ডায়নামিক্স — গ্রেডিয়েন্ট সংকট, এলএসটিএম সেল ও গেটেড মেমরি

রিকারেন্ট সিকোয়েন্স ডায়নামিক্স — গ্রেডিয়েন্ট সংকট, এলএসটিএম সেল ও গেটেড মেমরি

290 শব্দ • November 4, 2026 • ওয়েব সংস্করণ

❖



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

ভাষা, অডিও, স্টক মার্কেট বা বায়োফিজিক্যাল সিগন্যালের মতো ধারাবাহিক প্রক্রিয়ায় প্রতিটি মুহূর্তের ঘটনা তার পূর্বের দীর্ঘ ইতিহাসের ওপর নির্ভরশীল। **রিকারেন্ট নিউরাল নেটওয়ার্ক (RNN)** একটি লুকায়িত ভেক্টর h_t-এর মাধ্যমে এই অতীত স্মৃতি সংরক্ষণ করে। কিন্তু দীর্ঘ সিকোয়েন্সে চিরায়ত আরএনএন কুখ্যাত *ভ্যানিশিং ও এক্সপ্লোডিং গ্রেডিয়েন্ট সমস্যায়* আক্রান্ত হতো, যার ফলে অতীত স্মৃতি কয়েক ধাপ পরই মিলিয়ে যেত। ১৯৯৭ সালে সেপ হোচরামিটার এবং ইয়ুর্গেন শমিডহুবার **লং শর্ট-টার্ম মেমরি (LSTM)** উদ্ভাবনের মাধ্যমে এই ঐতিহাসিক সংকটের নিখুঁত সমাধান করেন।

১. ব্যাকপ্রোপাগেশন থ্রু টাইম (BPTT) ও সময়ের বিস্তার

একটি স্ট্যান্ডার্ড আরএনএন $h_t = sigma(W_{hh}h_{t-1} + W_{xh}x_t)$-এর মাধ্যমে অগ্রসর হয়। সময় বরাবর গ্রাফ উন্মোচন করে গ্রেডিয়েন্ট নির্ণয় করতে গেলে চেইন রুল প্রয়োগের ফলে জ্যাকোবিয়ান ম্যাট্রিক্সের গুণফল উৎপন্ন হয়:

$$fracpartialh_Tpartialh_t = prod_{k=t+1}^{T}textdiag(sigma'(z_k))W_{hh}^{T}$$

২. বেনজিওর স্পেকট্রাল ব্যাসার্ধ উপপাদ্য

১৯৯৪ সালে ইয়োশুয়া বেনজিও প্রমাণ করেন যে যদি ওজন ম্যাট্রিক্স W_{hh}-এর বৃহত্তম আইগেনভ্যালু বা স্পেকট্রাল ব্যাসার্ধ $rho(W_{hh})$ এবং অ্যাক্টিভেশন গ্রেডিয়েন্টের গুণফল ১-এর চেয়ে কম হয় (gamma rho < 1), তবে গ্রেডিয়েন্ট সময়ের দূরত্বের সাথে সাথে এক্সপোনেনশিয়াল হারে শূন্যে মিলিয়ে যায় ($|fracpartialh_Tpartialh_t|to0$)। ফলে ৫০ বা ১০০ ধাপ আগের কোনো তথ্যের প্রভাব মডেলে পৌঁছাতে পারে না।

Shakil Mahmud • Research Monograph

৩. এলএসটিএম (LSTM) এবং কনস্ট্যান্ট এরর ক্যারোজেল

হোচরামিটার ও শমিডহুবার আবিষ্কার করেন **Constant Error Carousel (CEC)**: তারা মেমরির জন্য একটি স্বতন্ত্র সেল স্টেট c_t তৈরি করেন যা ম্যাট্রিক্স গুণনের বদলে যোগের মাধ্যমে পরিচালিত হয়:

$$c_t = f_t odot c_{t-1} + i_t odot tilde{c}_t$$

যেখানে f_t হলো ফরগেট গেট, i_t হলো ইনপুট গেট এবং o_t হলো আউটপুট গেট। যোগভিত্তিক আপডেটের কারণে জ্যাকোবিয়ান ম্যাট্রিক্স গুণন ছাড়াই শত শত সময়-পদক্ষেপ ধরে গ্রেডিয়েন্ট অবিকল সংরক্ষিত থাকে।

৪. গেটেড রিকারেন্ট ইউনিট (GRU)

২০১৪ সালে কিয়াংহিউন চো এলএসটিএম-এর ফরগেট ও ইনপুট গেটকে একটি মাত্র আপডেট গেট z_t-এ একীভূত করে **জিআরইউ (GRU)** তৈরি করেন, যা ২৫% কম প্যারামিটারে সমমানের মেমরি সংরক্ষণ নিশ্চিত করে।



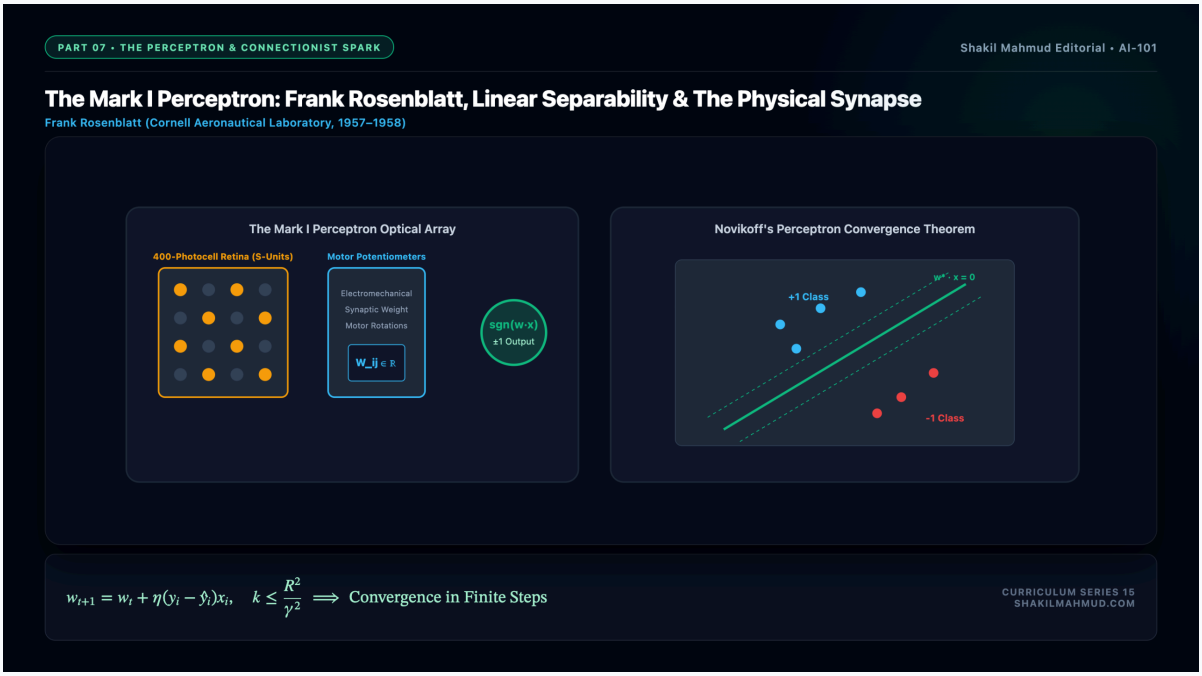
সাইটেশন: মাহমুদ, শাকিল (2026)। "রিকারেন্ট সিকোয়েন্স ডায়নামিক্স — গ্রেডিয়েন্ট সংকট, এলএসটিএম সেল ও গেটেড মেমরি"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ৬।
https://shakilmahmud.com/bn/deep-learning-architectures-part-06-recurrent-networks-lstm-gru-vanishing-gradients-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • সপ্তম খণ্ড

অ্যাটেনশন মেকানিজম ও কোয়েরি-কি-ভ্যালু দ্বৈততা — বাহদানাও থেকে স্কেল্ড ডট-প্রোডাক্ট

অ্যাটেনশন মেকানিজম ও কোয়েরি-কি-ভ্যালু দ্বৈততা — বাহদানাও থেকে স্কেল্ড ডট-প্রোডাক্ট

276 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



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

চিরায়ত সিকোয়েন্স-টু-সিকোয়েন্স মডেলে এনকোডারকে একটি সম্পূর্ণ বাক্যকে একটি মাত্র নির্দিষ্ট মাত্রার ভেক্টরে ($cinmathbbR^d$) সংকুচিত করতে বাধ্য করা হতো। বাক্যের দৈর্ঘ্য বৃদ্ধির সাথে সাথে এই সংকুচিত বটলনেকে তথ্য নষ্ট হয়ে যেত। ২০১৪ সালে জিমিত্রি বাহদানাও, কিয়াংহিউন চো এবং ইয়োশুয়া বেনজিও আবিষ্কার করেন **অ্যাটেনশন মেকানিজম (Attention Mechanism)**। স্থির ভেক্টরের বদলে গতিশীল মেমরি অ্যাড্রেসিং ব্যবহারের মাধ্যমে এটি নিউরাল নেটওয়ার্ককে বটলনেক থেকে মুক্ত করে এবং আধুনিক এআই বিপ্লবের ভিত রচনা করে।

১. এনকোডার বটলনেক সংকট

দীর্ঘ বাক্য অনুবাদে শেষ হিডেন স্টেট $c = h_T$ বাক্যের সমস্ত ব্যাকরণ ও অর্থ ধরে রাখতে পারত না। ২০ শব্দের বেশি হলেই মডেলের অনুবাদ মান ব্যাপকভাবে ধসে পড়ত।

২. বাহদানাও অ্যাডিটিভ অ্যাটেনশন

বাহদানাও প্রস্তাব করেন যে ডিকোডারের প্রতিটি পদক্ষেপে একটি গতিশীল কনটেক্সট ভেক্টর $c_i = sumalpha_{ij}h_j$ তৈরি করা হবে। অ্যালাইনমেন্ট এনার্জি $e_{ij} = v_a^T tanh(W_a s_{i-1} + U_a h_j)$-এর ওপর সফটম্যাক্স প্রয়োগ করে মডেল স্বয়ংক্রিয়ভাবে শেখে বাক্যের কোন অংশে মনোযোগ দেওয়া প্রয়োজন।

৩. কোয়েরি, কি এবং ভ্যালু (QKV) দ্বৈততা

২০১৭ সালে ভাসওয়ানি এবং তার দল অ্যাটেনশনকে কম্পিউটার সায়েন্সের মেমরি রিট্রিভাল হিসেবে সংজ্ঞায়িত করেন:

Shakil Mahmud • Research Monograph

- **কোয়েরি (Q):** বর্তমান অনুসন্ধান বা প্রশ্ন।
- **কি (K):** মেমরির অ্যাড্রেস বা সূচক।
- **ভ্যালু (V):** মেমরিতে সংরক্ষিত প্রকৃত তথ্য।

সাধারণ ডাটাবেজে একটি কোয়েরির সাথে একটি মাত্র কি মেলে। কিন্তু নিউরাল অ্যাটেনশন সমস্ত সংরক্ষিত তথ্যের ওপর একটি মসৃণ ও অন্তরীকরণযোগ্য সম্ভাব্যতা বণ্টন তৈরি করে তথ্য আহরণ করে।

৪. স্কেল্ড ডট-প্রোডাক্ট অ্যাটেনশনের গাণিতিক স্কেলিং

ম্যাট্রিক্স ডট প্রোডাক্ট QK^T দ্রুত হলেও এর ভ্যারিয়েন্স ডাইমেনশন d_k-এর সমান হয় ($textVar(q^Tk) = d_k$)। ফলে বড় মাত্রায় মান অনেক বেড়ে যায় এবং সফটম্যাক্স স্যাচুরেশনে চলে গিয়ে গ্রেডিয়েন্ট শূন্য হয়ে যায়। এই কারণে $sqrtd_k$ দিয়ে ভাগ করে ভ্যারিয়েন্স ১-এ নামিয়ে আনা হয়:

$$textAttention(Q, K, V) = textsoftmaxleft(fracQK^TsqrtdkrightV$$

যা আধুনিক ট্রান্সফরমার আর্কিটেকচারের কেন্দ্রবিন্দু।



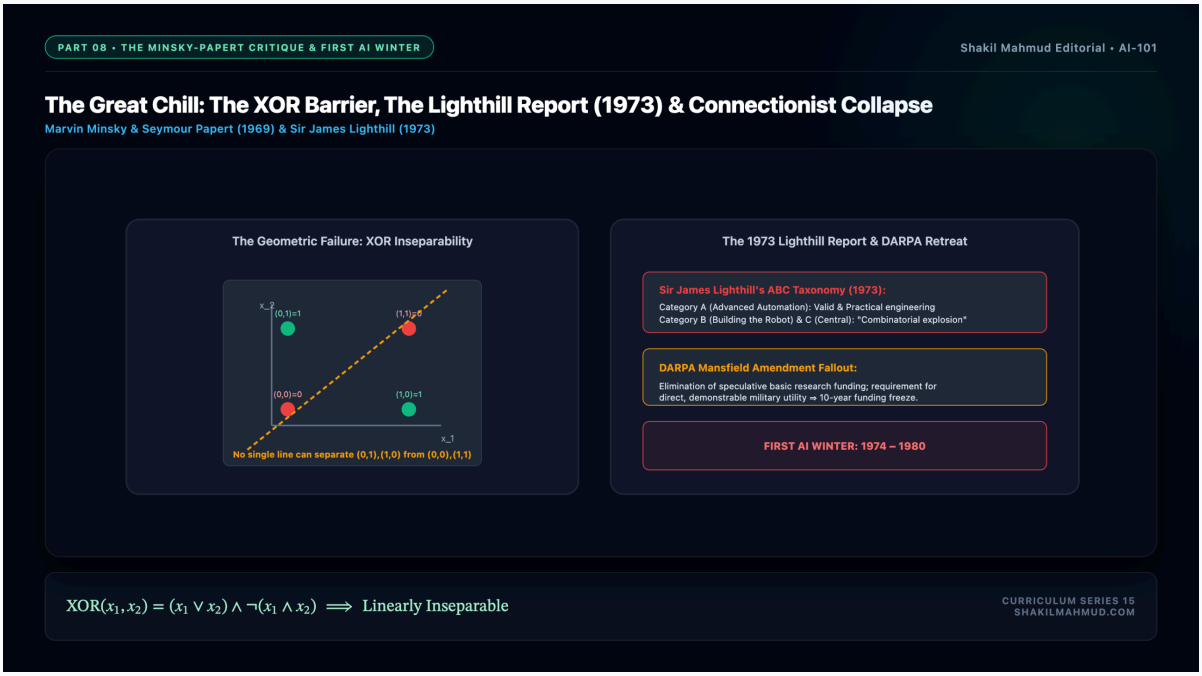
সাইটেশন: মাহমুদ, শাকিল (2026)। "অ্যাটেনশন মেকানিজম ও কোয়েরি-কি-ভ্যালু দ্বৈততা — বাহদানাউ থেকে স্কেল্ড ডট-প্রোডাক্ট"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ৭।
https://shakilmahmud.com/bn/deep-learning-architectures-part-07-attention-mechanisms-query-key-value-duality-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • অষ্টম খণ্ড

ট্রান্সফরমার আর্কিটেকচার — মাল্টি-হেড অ্যাটেনশন, পজিশনাল এনকোডিং ও রেসিডুয়াল স্ট্রিম

ট্রান্সফরমার আর্কিটেকচার — মাল্টি-হেড অ্যাটেনশন, পজিশনাল এনকোডিং ও রেসিডুয়াল স্ট্রিম

267 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।*

২০১৭ সালের জুনে প্রকাশিত আশিষ ভাসওয়ানি ও সহকর্মীদের যুগান্তকারী গবেষণাপত্র *“Attention Is All You Need”* উন্মোচন করে **ট্রান্সফরমার আর্কিটেকচার**। রিকারেন্স ও কনভোলিউশন বর্জন করে সম্পূর্ণভাবে সেলফ-অ্যাটেনশনের ওপর ভিত্তি করে গড়ে ওঠা এই মডেল আরএনএন-এর সিকোয়েন্সিয়াল সীমাবদ্ধতা চূর্ণ করে দেয়। এটি জিপিইউ হার্ডওয়্যারের পূর্ণ সমান্তরাল ব্যবহারের সুযোগ তৈরি করে এবং যেকোনো দুটি টোকেনের মধ্যবর্তী দূরত্বকে $O(1)$-এ নামিয়ে আনে। আজ ট্রান্সফরমার আধুনিক এআই ও ফাউন্ডেশন মডেলের অবিসংবাদিত ভিত্তি।

১. মাল্টি-হেড অ্যাটেনশন (MHA)

একটি মাত্র অ্যাটেনশন হেড ভাষার একাধিক ব্যাকরণিক ও অর্থগত দিক একসঙ্গে অনুধাবন করতে পারে না। ট্রান্সফরমার ইনপুটকে h সংখ্যক স্বাধীন সাবস্পেসে বিভক্ত করে:

$$texthead_i = textAttention(QW_i^Q, KW_i^K, VW_i^V)$$

$$textMultiHead(Q,K,V) = textConcat(texthead_1, dots, texthead_h)W^O$$

এর ফলে কোনো হেড বাক্যের ব্যাকরণিক ক্রিয়াকে অনুসরণ করে, কোনো হেড সর্বনামের নির্দেশককে চিহ্নিত করে, আবার কোনো হেড দীর্ঘ দূরত্বের ঐতিহাসিক সম্পর্ক ট্র্যাক করে।

Shakil Mahmud • Research Monograph

২. সাইনুসয়েডাল পজিশনাল এনকোডিং

সেলফ-অ্যাটেনশন নিজে থেকে শব্দের ক্রম বা অবস্থান বুঝতে পারে না (Permutation Equivariant)। এই সমস্যা সমাধানে ভাসওয়ানি সাইন ও কোসাইন তরঙ্গের সমন্বয়ে পজিশনাল ভেক্টর যোগ করেন:

$$textPE_{(pos,2i)} = sinleft(fracpos10000^{2i/d}right), quadtextPE_{(pos,2i+1)} = cosleft(fracpos10000^{2i/d}right)$$

ত্রিকোণমিতিক নিয়মের কারণে যেকোনো দুটি অবস্থানের আপেক্ষিক দূরত্ব একটি লিনিয়ার রূপান্তর হিসেবে প্রকাশ পায়।

৩. পজিশন-ওয়াইজ ফিড-ফরওয়ার্ড নেটওয়ার্ক (FFN)

অ্যাটেনশন কেবল বিদ্যমান টোকেন ভেক্টরগুলোর সংমিশ্রণ তৈরি করে, কিন্তু নতুন কোনো নন-লিনিয়ার বৈশিষ্ট্য তৈরি করতে পারে না। প্রতিটি টোকেনে স্বতন্ত্রভাবে ৪ গুণ সম্প্রসারিত FFN প্রয়োগ করে নন-লিনিয়ার প্রজ্ঞা সঞ্চার করা হয়।

৪. রেসিডুয়াল স্ট্রিম: কেন্দ্রীয় যোগাযোগ বাস

আধুনিক মেকানিস্টিক ইন্টারপ্রেটেবিলিটি অনুযায়ী ট্রান্সফরমার হলো একটি অবিচ্ছিন্ন **রেসিডুয়াল স্ট্রিম (Residual Stream)**। প্রতিটি স্তর এই কেন্দ্রীয় বাস থেকে তথ্য পাঠ করে এবং প্রক্রিয়াজাত ফলাফল যোগ আকারে বাসে লিখে দেয়, যা বৃহৎ মডেল স্কেলিংয়ের মূল রহস্য।

সাইটেশন: মাহমুদ, শাকিল (2026)। "ট্রান্সফরমার আর্কিটেকচার — মাল্টি-হেড অ্যাটেনশন, পজিশনাল এনকোডিং ও রেসিডুয়াল স্ট্রিম"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ৮।
https://shakilmahmud.com/bn/deep-learning-architectures-part-08-transformer-architecture-multihead-attention-residual-stream-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • নবম খণ্ড

রোটারি পজিশন এম্বেডিং (RoPE) ও কনটেক্সট উইন্ডো — জটিল ঘূর্ণন ও দীর্ঘ দূরত্বের বিস্তার

রোটারি পজিশন এম্বেডিং (RoPE) ও কনটেক্সট উইন্ডো — জটিল ঘূর্ণন ও দীর্ঘ দূরত্বের বিস্তার

319 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



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

আদি ট্রান্সফরমার মডেলে ইনপুট ভেক্টরের সাথে সরাসরি সাইন ও কোসাইন পজিশন ভেক্টর যোগ করা হতো ($x_i + p_i$)। কিন্তু এই যোগভিত্তিক অবস্থান নির্দেশক মডেলকে তার প্রশিক্ষণ দৈর্ঘ্যের বাইরে দীর্ঘ টেক্সট পড়তে মারাত্মকভাবে বাধা দেয়। ২০২১ সালে জিয়ানলিন সু আবিষ্কার করেন **রোটারি পজিশন এম্বেডিং (RoPE)**। এটি কোনো ভেক্টর যোগ না করে, জটিল সমতলে (Complex Plane) অর্থোগোনাল ঘূর্ণন ম্যাট্রিক্সের সাহায্যে টোকেনের আপেক্ষিক দূরত্বকে নিখুঁত জ্যামিতিক কোণে রূপান্তরিত করে। আজ LLaMA, Mistral, Qwen ও DeepSeek-এর মতো বিশ্বের শীর্ষ মডেলগুলো RoPE ব্যবহার করে।

১. আপেক্ষিক অবস্থানের গাণিতিক উদ্দেশ্য

অবস্থান m এবং n-এ থাকা দুটি টোকেনের কোয়েরি q_m এবং কি k_n-এর ইনার প্রোডাক্ট কেবল তাদের শব্দার্থ এবং তাদের মধ্যবর্তী আপেক্ষিক ব্যবধান m – n-এর ওপর নির্ভর করা উচিত:

langle f_q(x_m, m), f_k(x_n, n) rangle = g(x_m, x_n, m – n)

চিরায়ত পদ্ধতিতে পজিশন ভেক্টর যোগ করায় বিভিন্ন ক্রস-টার্ম তৈরি হয়ে আপেক্ষিক দূরত্ব বিকৃত হয়ে যেত।

২. দ্বিমাত্রিক জটিল ঘূর্ণনের উপপাদন

জটিল সংখ্যা পদ্ধতিতে ভেক্টর দুটিকে কোণ $mtheta$ ও $ntheta$ দ্বারা আবর্তন করানো হয়:

Shakil Mahmud • Research Monograph

$$q_m = qe^{imtheta}, quadk_n = ke^{intheta}$$

হারমিশিয়ান ইনার প্রোডাক্ট নিলে দেখা যায়:

langle q_m, k_n rangle = text{Re}[(q e^{i m theta}) (k e^{i n theta})^*] = text{Re}[q k^* e^{i (m – n)theta}]

এখানে পরম অবস্থান m এবং n সম্পূর্ণ বিলুপ্ত হয়ে যায় এবং অবশিষ্ট থাকে কেবল আপেক্ষিক দূরত্বের কোণ (m – n)theta।

৩. উচ্চ মাত্রায় ব্লক-ডায়াগোনাল ঘূর্ণন

d-মাত্রিক স্পেসকে $d/2$ সংখ্যক দ্বিমাত্রিক অর্থোগোনাল সমতলে বিভক্ত করে প্রতিটি সমতলের জন্য পৃথক ফ্রিকোয়েন্সি $theta_j = 10000^{-2(j-1)/d}$ নির্ধারণ করা হয়। এটি কোনো ম্যাট্রিক্স গুণন ছাড়াই ভেক্টর এলিমেন্ট-ওয়াইজ অপারেশনের মাধ্যমে $O(d)$ সময়ে হিসাব করা যায়।

৪. দীর্ঘ কনটেক্সট উইন্ডো সম্প্রসারণ: NTK ও YaRN

মডেলকে ৪ হাজার টোকেন থেকে ১ লাখ ২৮ হাজার টোকেনে সম্প্রসারণ করতে **NTK-Aware Scaling** এবং **YaRN** উদ্ভাবন করা হয়। উচ্চ ফ্রিকোয়েন্সি (যা স্থানীয় ব্যাকরণ বোঝায়) অবিকৃত রেখে নিম্ন ফ্রিকোয়েন্সিগুলোকে (যা দূরবর্তী ক্রম বোঝায়) সূক্ষ্মভাবে স্কেল করার মাধ্যমে মডেল কোনো অতীত তথ্য না ভুলেই অসীম দৈর্ঘ্যের টেক্সট অনুধাবন করতে সক্ষম হয়।

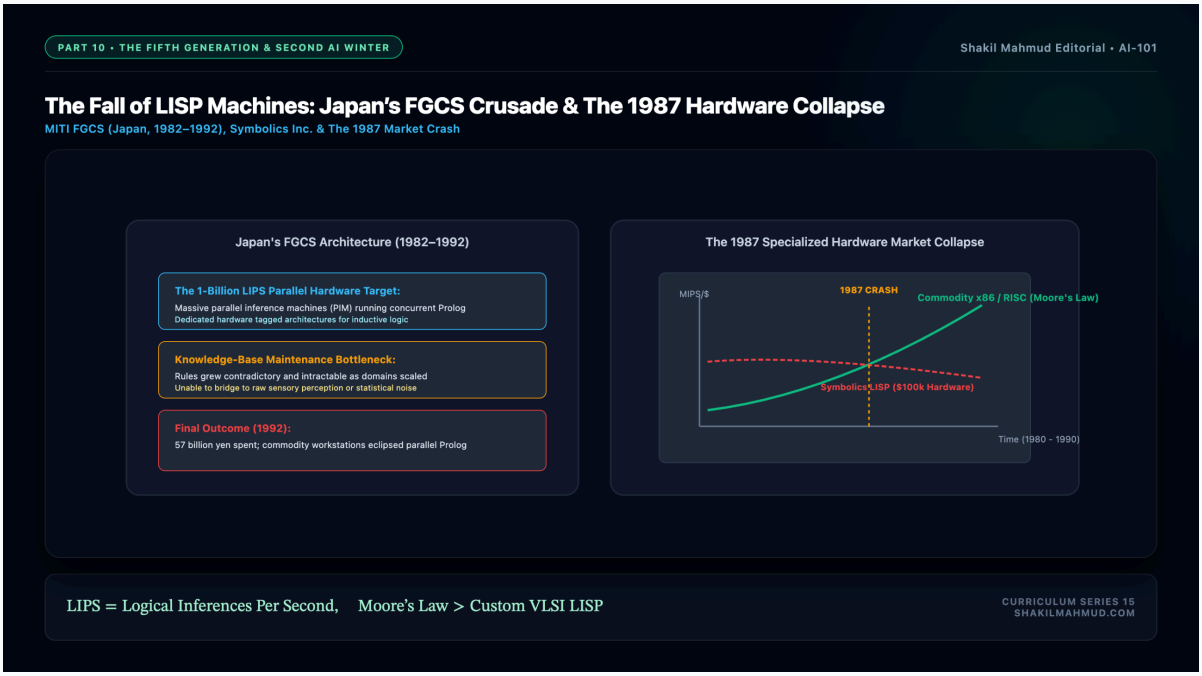
সাইটেশন: মাহমুদ, শাকিল (2026)। "রোটারি পজিশন এম্বেডিং (RoPE) ও কনটেক্সট উইন্ডো — জটিল ঘূর্ণন ও দীর্ঘ দূরত্বের বিস্তার"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ৯।
https://shakilmahmud.com/bn/deep-learning-architectures-part-09-rotary-position-embeddings-rope-context-windows-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • দশম খণ্ড

অটোএনকোডার ও ল্যাটেন্ট রিপ্রেজেন্টেশন — ম্যানিফোল্ড বটলনেক ও ডিসএন্ট্যাঙ্গলমেন্ট

অটোএনকোডার ও ল্যাটেন্ট রিপ্রেজেন্টেশন — ম্যানিফোল্ড বটলনেক ও ডিসএন্ট্যাঙ্গলমেন্ট

258 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।*

আনসুপারভাইজড লার্নিংয়ের অন্যতম প্রধান উদ্দেশ্য হলো কোনো লেবেল ছাড়াই জটিল উচ্চ-মাত্রিক ডেটার অন্তর্নিহিত চালিকাশক্তিগুলোকে একটি সংক্ষিপ্ত গাণিতিক রূপ দেওয়া। **অটোএনকোডার (Autoencoders)** ইনপুট ডেটা $xinmathbbR^d$-কে একটি সংকীর্ণ বটলনেক ভেক্টর $zinmathbbR^k$ ($klld$)-এর মধ্য দিয়ে সংকুচিত করে পুনরায় হুবহু ইনপুট তৈরি করার চেষ্টা করে। এই বাধ্যবাধকতার কারণে মডেল ডেটার অপ্রয়োজনীয় নয়েজ ফেলে দিয়ে মূল অন্তর্নিহিত **লো-ডাইমেনশনাল ম্যানিফোল্ড** শনাক্ত করতে শেখে।

১. আন্ডারকমপ্লিট অটোএনকোডারের মূলনীতি

একটি অটোএনকোডার দুটি প্রধান অংশে বিভক্ত:

- **এনকোডার ($fphi$):** ইনপুটকে ল্যাটেন্ট কোডে সংকুচিত করে: $z = sigma(W_ex + b_e)$।
- **ডিকোডারের ($gpsi$):** ল্যাটেন্ট কোড থেকে মূল ইনপুট পুনর্গঠন করে: $hatx = sigma(W_dz + b_d)$।

যদি অ্যাক্টিভেশন ফাংশন সম্পূর্ণ লিনিয়ার হয়, তবে অটোএনকোডার হুবহু ক্লাসিক্যাল পিসিএ (PCA)-এর সমতুল্য হয়। নন-লিনিয়ার স্তরের কারণে এটি বাঁকানো ম্যানিফোল্ডের জ্যামিতি নিখুঁতভাবে আয়ত্ত করতে পারে।

Shakil Mahmud • Research Monograph

২. ডিনয়েজিং অটোএনকোডার (DAE) ও স্কোর ম্যাচিং

ইনপুটকে অবিকল কপি করার ফাঁকি রোধ করতে ২০০৮ সালে পাস্কাল ভিনসেন্ট উদ্ভাবন করেন **ডিনয়েজিং অটোএনকোডার**। ইনপুটে কৃত্রিম নয়েজ যোগ করে মডেলকে মূল পরিষ্কার ডেটা পুনর্গঠনের নির্দেশ দেওয়া হয়। ২০১১ সালে বেনজিও প্রমাণ করেন যে এই প্রক্রিয়াটি আসলে ডেটার সম্ভাব্যতা ঘনত্বের গ্রেডিয়েন্ট বা **স্কোর ফাংশন (**$nablalogp(x)$**)** নির্ণয় করে, যা আধুনিক ডিফিউশন মডেলের মূল তাত্ত্বিক ভিত্তি।

৩. কন্ট্রাক্টিভ অটোএনকোডার (CAE)

কন্ট্রাক্টিভ অটোএনকোডারে জ্যাকোবিয়ান ম্যাট্রিক্সের ফ্রবেনিয়াস নর্ম ($|J_f|_F^2$) যুক্ত করে শাস্তি দেওয়া হয়। এর ফলে ডেটা ম্যানিফোল্ডের স্পর্শক বরাবর তথ্য সংরক্ষিত থাকে, কিন্তু লম্ব দিকের সমস্ত ক্ষুদ্র বিচ্যুতি ও নয়েজ সম্পূর্ণরূপে ধ্বংস হয়ে যায়।

৪. রিপ্রেজেন্টেশন ডিসএন্ট্যাঙ্গলমেন্ট

ল্যাটেন্ট কোডের প্রতিটি একক মাত্রা যদি বাস্তব জগতের একটি মাত্র স্বাধীন বৈশিষ্ট্যের (যেমন বস্তুর আকার, কোণ বা আলোর তীব্রতা) সাথে সরাসরি সম্পর্কযুক্ত হয়, তবে তাকে ডিসএন্ট্যাঙ্গলড রিপ্রেজেন্টেশন বলে।

সাইটেশন: মাহমুদ, শাকিল (2026)। "অটোএনকোডার ও ল্যাটেন্ট রিপ্রেজেন্টেশন — ম্যানিফোল্ড বটলনেক ও ডিসএন্ট্যাঙ্গলমেন্ট"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১০।
https://shakilmahmud.com/bn/deep-learning-architectures-part-10-autoencoders-latent-representations-manifold-disentanglement-bn/

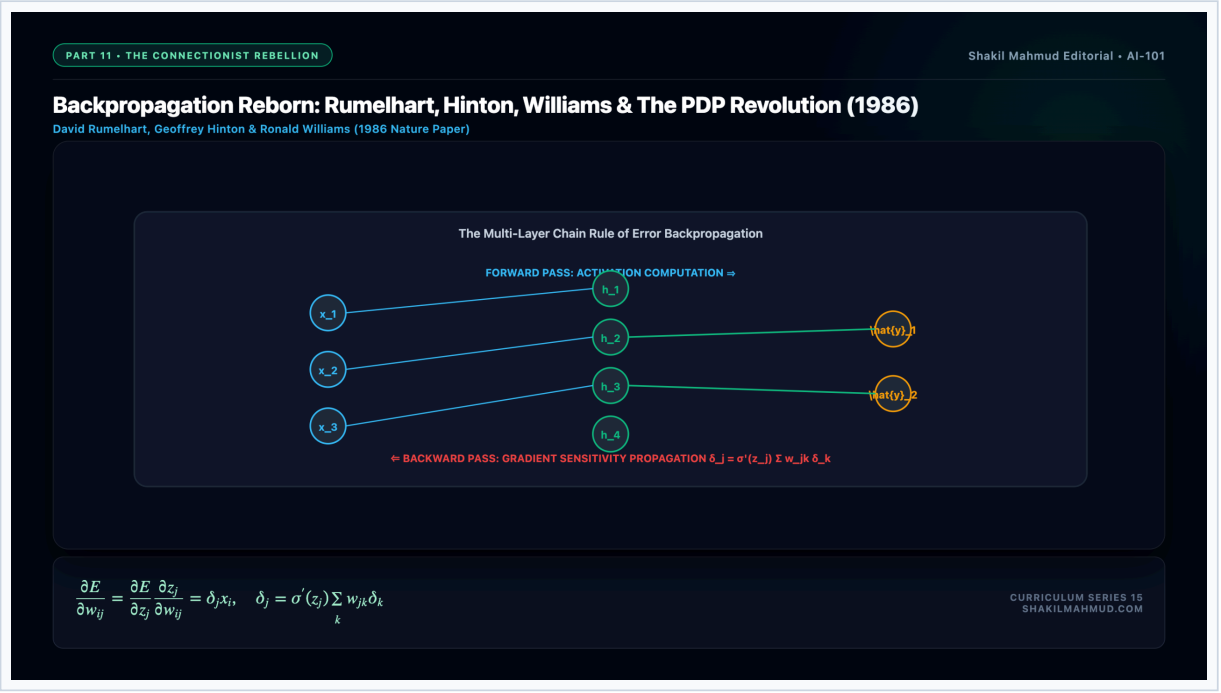
শাকিল মাহমুদ মনোগ্রাফ সিরিজ • একাদশ খণ্ড

ভ্যারিয়েশনাল অটোএনকোডার (VAE) — এভিডেন্স লোয়ার বাউন্ড ও রিপ্যারামিটারাইজেশন ট্রিক

ভ্যারিয়েশনাল অটোএনকোডার (VAE) — এভিডেন্স লোয়ার বাউন্ড ও রিপ্যারামিটারাইজেশন ট্রিক

263 শব্দ • November 4, 2026 • ওয়েব সংস্করণ

❖



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

চিরায়ত ডিটারমিনিস্টিক অটোএনকোডারের একটি প্রধান দুর্বলতা হলো—এর ল্যাটেন্ট স্পেস মসৃণ হয় না; সেখানে ফাঁকা ও বিশৃঙ্খল অঞ্চল তৈরি হয় যেখান থেকে নতুন কোনো অর্থপূর্ণ ছবি বা নমুনা তৈরি করা যায় না। ২০১৩ সালে ডিডেরিক কিংমা এবং ম্যাক্স ওয়েলিং উদ্ভাবন করেন **ভ্যারিয়েশনাল অটোএনকোডার (VAE)**। বায়েসিয়ান ভ্যারিয়েশনাল ইনফারেন্সের ওপর প্রতিষ্ঠিত এই মডেলটি স্টোক্যাস্টিক র‍্যান্ডম স্যাম্পলিংয়ের মধ্য দিয়েও গ্রেডিয়েন্ট প্রবাহিত করার জন্য বিখ্যাত **রিপ্যারামিটারাইজেশন ট্রিক (Reparameterization Trick)** উপহার দেয়।

১. মার্জিনাল লাইকলিহুড ও ইন্ট্র্যাক্টেবিলিটি

ল্যাটেন্ট ভ্যারিয়েবল মডেল $xsimp_theta(xmidz)$-এ মার্জিনাল সম্ভাব্যতা $logp(x) = logintp(xmidz)p(z)dz$ হিসাব করা কোনো গভীর নেটওয়ার্কের জন্য বাস্তবসম্মত নয়। এর ফলে প্রকৃত পোস্টেরিয়র $p(zmidx)$-ও অপ্রাপ্য থেকে যায়।

২. এভিডেন্স লোয়ার বাউন্ড (ELBO)-এর উপপাদন

একটি এনকোডার নেটওয়ার্ক $q_phi(zmidx)$ দিয়ে প্রকৃত পোস্টেরিয়রকে অনুমান করা হয়। কুলব্যাক-লেইবলার (KL) ডাইভারজেন্সের সমীকরণ থেকে পাওয়া যায় বিখ্যাত **ELBO** সমীকরণ:

log p_theta(x) ge mathcal{L}_{text{ELBO}} = mathbb{E}_{q_phi(z mid x)}[log p_theta(x mid z)] – D_{text{KL}}(q_phi(z mid x) parallel p(z))

Shakil Mahmud • Research Monograph

এই সমীকরণের দুটি অংশ রয়েছে:

1. **পুনর্গঠন যথার্থতা (Reconstruction):** ডিকোডার যাতে ল্যাটেন্ট কোড থেকে ইনপুটকে নিখুঁতভাবে পুনর্নির্মাণ করতে পারে।
2. **ল্যাটেন্ট নিয়মিতকরণ (KL Divergence):** ল্যাটেন্ট স্পেসের বিন্যাস যেন আদর্শ গাউসিয়ান $mathcalN(0, I)$-এর কাছাকাছি থাকে, যাতে কোনো বিচ্ছিন্ন ফাঁকা অঞ্চল না থাকে।

৩. রিপ্যারামিটারাইজেশন ট্রিক (Reparameterization Trick)

র‍্যান্ডম স্যাম্পলিং নোডের মধ্য দিয়ে সরাসরি ব্যাকপ্রোপাগেশন চালানো অসম্ভব, কারণ স্যাম্পলিংয়ের কোনো অন্তরক (Derivative) নেই। কিংমা ও ওয়েলিং স্টোক্যাস্টিক নয়েজকে একটি স্বাধীন চলক $epsilonsimmathcalN(0, I)$ হিসেবে আলাদা করে দেন:

$$z = mu(x) + sigma(x)odotepsilon$$

এই সহজ কিন্তু যুগান্তকারী সমীকরণের ফলে র‍্যান্ডম স্যাম্পলিং একটি সাধারণ স্কেলিং ও যোগ অপারেশনে রূপান্তরিত হয়, যার মধ্য দিয়ে ব্যাকপ্রোপাগেশনের চেইন রুল অবিকল প্রবাহিত হতে পারে।

সাইটেশন: মাহমুদ, শাকিল (2026)। "ভ্যারিয়েশনাল অটোএনকোডার (VAE) — এভিডেন্স লোয়ার বাউন্ড ও রিপ্যারামিটারাইজেশন ট্রিক"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১১।
https://shakilmahmud.com/bn/deep-learning-architectures-part-11-variational-autoencoders-elbo-reparameterization-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • দ্বাদশ খণ্ড

জেনারেটিভ অ্যাডভারসারিয়াল নেটওয়ার্ক (GAN) — মিনিম্যাক্স গেম ও ওয়াসারস্টাইন জ্যামিতি

জেনারেটিভ অ্যাডভারসারিয়াল নেটওয়ার্ক (GAN) — মিনিম্যাক্স গেম ও ওয়াসারস্টাইন জ্যামিতি

285 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



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

২০১৪ সালে ইয়ান গুডফেলো সম্পূর্ণ ভিন্নধর্মী এক জেনারেটিভ দর্শনের অবতারণা করেন যা কোনো সম্ভাব্যতার অনুমান ছাড়াই কাজ করে: **জেনারেটিভ অ্যাডভারসারিয়াল নেটওয়ার্ক (GAN)**। একে তুলনা করা যায় একজন জাল নোট প্রস্তুতকারক (Generator) এবং একজন পুলিশ গোয়েন্দার (Discriminator) মধ্যকার দ্বৈত লড়াইয়ের সাথে। ডিসক্রিমিনেটর আসল ও নকলের পার্থক্য ধরার চেষ্টা করে, আর জেনারেটর তাকে ফাঁকি দেওয়ার মতো নিখুঁত নমুনা তৈরি করে। পরবর্তীতে অপ্টিমাল ট্রান্সপোর্ট থিওরির **ওয়াসারস্টাইন দূরত্ব (Wasserstein Distance)** যুক্ত হয়ে এই প্রশিক্ষণকে চরম স্থিতিশীলতায় রূপ দেয়।

১. মিনিম্যাক্স জিরো-সাম গেম

মডেলের সামগ্রিক উদ্দেশ্য নির্ধারিত হয় $V(D, G)$ ফাংশনটির মাধ্যমে:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$$

স্থির জেনারেটরের জন্য সর্বোত্তম ডিসক্রিমিনেটর হলো $D^*(x) = frac{p_{\text{data}}}(x)}{p_{\text{data}}(x) + p_g(x)}$। এটিকে মূল সমীকরণে বসালে দেখা যায় যে জেনারেটর আসলে বাস্তব উপাত্ত ও কৃত্রিম উপাত্তের **জেনসেন-শ্যানন ডাইভারজেন্স (Jensen-Shannon Divergence)** সর্বনিম্ন করার চেষ্টা করছে।

Shakil Mahmud • Research Monograph

২. মোড কোলাপ্স ও ভ্যানিশিং গ্রেডিয়েন্ট সমস্যা

বাস্তব জগতের ডেটা উচ্চ-মাত্রিক স্পেসের একটি অতি ক্ষুদ্র ম্যানিফোল্ডে অবস্থান করে। ফলে আসল ও নকল ডেটার ম্যানিফোল্ড সাধারণত একে অপরকে স্পর্শই করে না। এই অবস্থায় জেনসেন-শ্যানন ডাইভারজেন্স একটি ধ্রুবক মানে ($log 2$) স্থির হয়ে যায় এবং তার গ্রেডিয়েন্ট শূন্য হয়ে পড়ে। এর ফলে ডিসক্রিমিনেটর সহজেই জিতে যায় এবং জেনারেটর কোনো শেখার সংকেত পায় না। অনেক সময় জেনারেটর কেবল একটি মাত্র সহজ ছবি বারবার তৈরি করে ডিসক্রিমিনেটরকে ফাঁকি দেওয়ার চেষ্টা করে, যাকে **মোড কোলাপ্স (Mode Collapse)** বলে।

৩. ওয়াসারস্টাইন গ্যান (WGAN) ও অপ্টিমাল ট্রান্সপোর্ট

২০১৭ সালে মার্টিন আরজভস্কি জেএস ডাইভারজেন্সের বদলে **আর্থ-মুভার দূরত্ব (Wasserstein-1 Distance)** প্রয়োগ করেন:

$$W_1(p_{\text{data}}, p_g) = \inf_{\gamma} \mathbb{E}[|x – y|]$$

ক্যান্টোরোভিচ-রুবিনস্টাইন দ্বৈততা উপপাদ্যের সাহায্যে ডিসক্রিমিনেটরকে ১-লিপশিটজ ক্রুসেডারে রূপান্তর করা হয়। ওয়াসারস্টাইন দূরত্বের সবচেয়ে বড় সৌন্দর্য হলো—উপাত্তের ম্যানিফোল্ড দুটি সম্পূর্ণ আলাদা হলেও এটি মসৃণ ও অর্থপূর্ণ দূরত্ব নির্দেশ করে, যা গ্রেডিয়েন্ট প্রবাহকে সর্বদা সচল রাখে এবং মোড কোলাপ্স দূর করে।

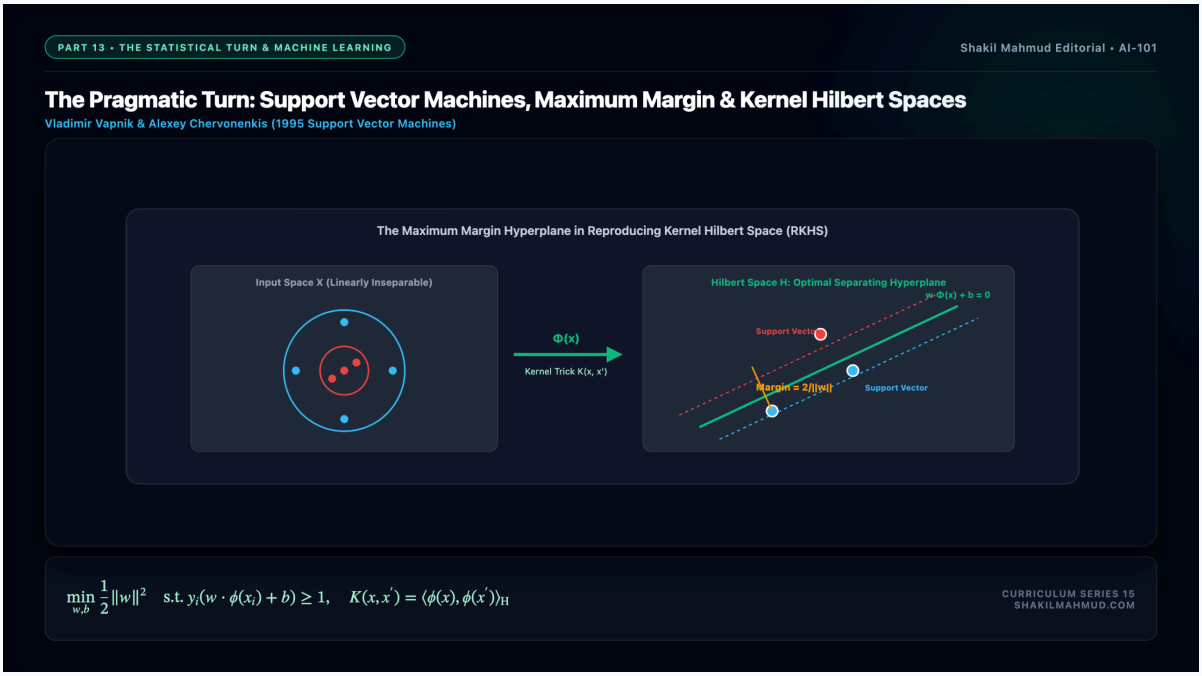
সাইটেশন: মাহমুদ, শাকিল (2026)। "জেনারেটিভ অ্যাডভারসারিয়াল নেটওয়ার্ক (GAN) — মিনিম্যাক্স গেম ও ওয়াসারস্টাইন জ্যামিতি"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১২।
https://shakilmahmud.com/bn/deep-learning-architectures-part-12-generative-adversarial-networks-gans-wasserstein-distance-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • ত্রয়োদশ খণ্ড

নরমালাইজিং ফ্লো ও উভমুখী নিউরাল নেটওয়ার্ক — চলক রূপান্তর ও নিখুঁত সম্ভাবনা

নরমালাইজিং ফ্লো ও উভমুখী নিউরাল নেটওয়ার্ক — চলক রূপান্তর ও নিখুঁত সম্ভাবনা

287 শব্দ • November 4, 2026 • ওয়েব সংস্করণ



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

ভিএই (VAE) বা গ্যান (GAN)-এর মতো প্রচলিত জেনারেটিভ মডেলগুলোর কিছু মৌলিক সীমাবদ্ধতা রয়েছে: ভিএই কেবল সম্ভাব্যতার একটি নিম্নসীমা (ELBO) অপ্টিমাইজ করতে পারে, আর গ্যান কোনো সম্ভাব্যতা বণ্টনই দিতে পারে না। **নরমালাইজিং ফ্লো (Normalizing Flows)** এই সংকটের সমাধান করে এক অনন্য গাণিতিক কাঠামোর মাধ্যমে—তারা তৈরি করে *উভমুখী বা বাইজেক্টিভ (Bijective) নিউরাল নেটওয়ার্ক*। বহু-চলক ক্যালকুলাসের **চলক রূপান্তর উপপাদ্য (Change-of-Variables Theorem)** ব্যবহার করে এটি একটি সাধারণ গাউসিয়ান বণ্টনকে বাস্তব জগতের অতি জটিল ডেটা বিন্যাসে রূপান্তর করে এবং একই সাথে **নিখুঁত মার্জিনাল লাইকলিহুড** প্রদান করে।

১. বহু-চলক চলক রূপান্তর উপপাদ্য

ধরি $ZinmathbbR^d$ হলো একটি আদর্শ গাউসিয়ান র‍্যান্ডম ভেক্টর যার ঘনত্ব $p_Z(z)$ সুপরিচিত। একটি মসৃণ ও উভমুখী ম্যাপিং f-এর মাধ্যমে যদি $X = f(Z)$ এবং $Z = f^{-1}(X)$ হয়, তবে সম্ভাব্যতা ভরের সংরক্ষণ নীতি অনুযায়ী:

$$p_X(x) = p_Z(f^{-1}(x)) cdot left|detleft(fracpartial f^{-1}(x)partial xright)right|$$

লগারিদম নিলে পাওয়া যায় নিখুঁত সম্ভাব্যতা সমীকরণ:

log p_X(x) = log p_Z(f^{-1}(x)) – log left| det J_f(f^{-1}(x)) right|

Shakil Mahmud • Research Monograph

যেখানে J_f হলো জ্যাকোবিয়ান ম্যাট্রিক্স।

২. জ্যাকোবিয়ান গণনা ও কাপলিং লেয়ার

একটি সাধারণ $dtimesd$ ম্যাট্রিক্সের নির্ণায়ক (Determinant) বের করতে $O(d^3)$ সময় লাগে, যা লক্ষ লক্ষ পিক্সেলে সম্পূর্ণ অকার্যকর। এই জটিলতা কাটিয়ে উঠতে এমন আর্কিটেকচার তৈরি করতে হয় যার জ্যাকোবিয়ান একটি **ট্রায়াঙ্গুলার ম্যাট্রিক্স**। ট্রায়াঙ্গুলার ম্যাট্রিক্সের নির্ণায়ক কেবল তার প্রধান কর্ণের উপাদানগুলোর গুণফলের সমান ($O(d)$)।

২০১৬ সালে ডিন এবং সহকর্মীরা উদ্ভাবন করেন **RealNVP**। তারা ইনপুটকে দুই ভাগে ভাগ করেন—এক ভাগকে অপরিবর্তিত রেখে তার ওপর ভিত্তি করে অপর ভাগে স্কেলিং ও ট্রানস্লেশন নেটওয়ার্ক প্রয়োগ করেন। এর ফলে জ্যাকোবিয়ান ম্যাট্রিক্সটি স্বয়ংক্রিয়ভাবে ট্রায়াঙ্গুলার হয় এবং নির্ণায়ক নির্ণয় অত্যন্ত সহজ হয়ে যায়।

৩. কন্টিনিউয়াস নরমালাইজিং ফ্লো (CNF)

স্তরের সংখ্যা অসীম সীমায় পৌঁছালে ফ্লো একটি সাধারণ ব্যবকলনীয় সমীকরণে (ODE) রূপান্তরিত হয়। লিউভিলের উপপাদ্য অনুযায়ী তখন নির্ণায়কের বদলে জ্যাকোবিয়ানের **ট্রেস (Trace)** ব্যবহৃত হয়, যা মেমরি ও সময়ের ব্যবহারে বিপ্লব ঘটায়।

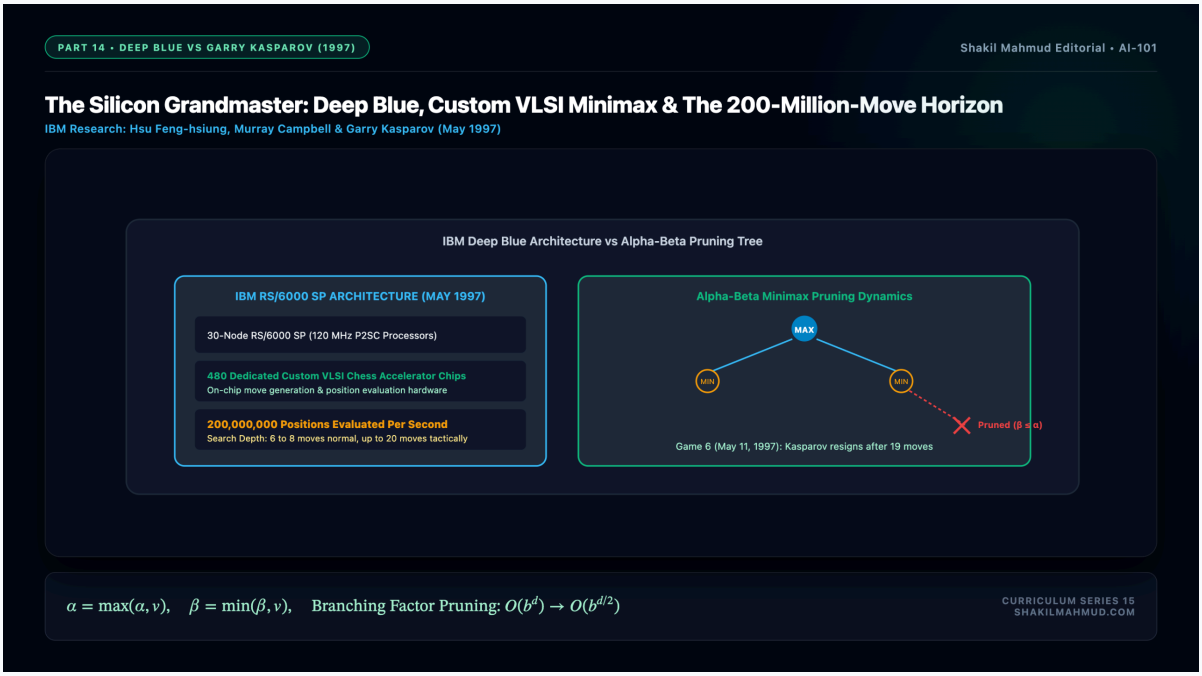
সাইটেশন: মাহমুদ, শাকিল (2026)। "নরমালাইজিং ফ্লো ও উভমুখী নিউরাল নেটওয়ার্ক — চলক রূপান্তর ও নিখুঁত সম্ভাবনা"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১৩।
https://shakilmahmud.com/bn/deep-learning-architectures-part-13-normalizing-flows-invertible-networks-change-of-variables-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • চতুর্দশ খণ্ড

এনার্জি-ভিত্তিক মডেল ও আধুনিক হপফিল্ড নেটওয়ার্ক — বোল্টজম্যান মেশিন থেকে ডেনস মেমরি

এনার্জি-ভিত্তিক মডেল ও আধুনিক হপফিল্ড নেটওয়ার্ক — বোল্টজম্যান মেশিন থেকে ডেনস মেমরি

310 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।*

অধিকাংশ নিউরাল মডেল আউটপুটে সফটম্যাক্স প্রয়োগ করে সম্ভাব্যতা ১-এ স্বাভাবিক (Normalize) করতে বাধ্য হয়। কিন্তু অতি-উচ্চ মাত্রার স্পেসে এই সাধারণীকরণ গণনা করা প্রায় অসম্ভব। ইয়ান লেকান এবং জিওফ্রে হিন্টনের নেতৃত্বাধীন **এনার্জি-ভিত্তিক মডেল (Energy-Based Models – EBM)** এই স্বাভাবিকীকরণের শর্ত প্রত্যাখ্যান করে। তারা একটি অবিচ্ছিন্ন স্কেলার এনার্জি ফাংশন $E_theta(x)$ সংজ্ঞায়িত করে—যেখানে বাস্তব ও যুক্তিসঙ্গত উপাত্তগুলো গভীর এনার্জি উপত্যকায় (Attractor Basin) অবস্থান করে এবং অবাস্তব উপাত্তগুলো উচ্চ এনার্জি বাধার সম্মুখীন হয়।

১. বোল্টজম্যান-গিবস বিন্যাস ও পার্টিশন ফাংশন

একটি এনার্জি ফাংশনের সাপেক্ষে সম্ভাব্যতা নির্ধারিত হয় বোল্টজম্যান সমীকরণ দ্বারা:

$$p_theta(x) = fracexp(-E_theta(x))Z(theta), quad Z(theta) = intexp(-E_theta(tildex))dtildex$$

এখানে হর $Z(theta)$-কে বলা হয় **পার্টিশন ফাংশন (Partition Function)**। উচ্চ মাত্রায় এই ইন্টিগ্রাল নির্ণয় অসম্ভব হলেও দুটি বিন্দুর আপেক্ষিক সম্ভাব্যতা পরিমাপের জন্য কেবল তাদের এনার্জির পার্থক্য E(x_2) – E(x_1) জানাই যথেষ্ট।

২. পজিটিভ বনাম নেগেটিভ ফেজ ডায়নামিক্স

লগ-লাইকলিহুডের গ্রেডিয়েন্ট হিসাব করলে প্রশিক্ষণ প্রক্রিয়ার দ্বৈত তাপগতীয় রূপ প্রকাশিত হয়:

Shakil Mahmud • Research Monograph

nabla_theta mathcal{L} = mathbb{E}_{x sim p_{text{data}}}[nabla_theta E_theta(x)] – mathbb{E}_{tilde{x} sim p_theta}[nabla_theta E_theta(tilde{x})]

- **পজিটিভ ফেজ:** বাস্তব উপাত্তের বিন্দুগুলোতে এনার্জি হ্রাস করে গভীর কুয়ো বা খাদ তৈরি করে।
- **নেগেটিভ ফেজ:** মডেলের কল্পিত বা ভুল নমুনাগুলোতে এনার্জি বৃদ্ধি করে চূড়া তৈরি করে।

৩. ল্যাঞ্জেভিন স্যাম্পলিং ও কনট্রাস্টিভ ডাইভারজেন্স

নেগেটিভ ফেজের নমুনা সংগ্রহে **ল্যাঞ্জেভিন এমসিএমসি (Langevin MCMC)** প্রয়োগ করা হয়। হিন্টন উদ্ভাবিত *কনট্রাস্টিভ ডাইভারজেন্স (CD-k)* বাস্তব ডেটা থেকে মার্কভ চেইন শুরু করে মাত্র এক বা কয়েক ধাপে কার্যকর নমুনা তৈরি করতে পারে।

৪. আধুনিক হপফিল্ড নেটওয়ার্ক ও ট্রান্সফরমার সমতুল্যতা

১৯৮২ সালে জন হপফিল্ড প্রবর্তিত ক্লাসিক্যাল হপফিল্ড নেটওয়ার্কের মেমরি ধারণক্ষমতা ছিল মাত্র $Capprox0.14d$। ২০১৬ সালে ক্রটভ ও হপফিল্ড এবং ২০২০ সালে রামসাওয়ার এক্সপোনেনশিয়াল এনার্জি মিথস্ক্রিয়া প্রবর্তন করেন:

$$E(x) = -textlse(betaX^Tx) + frac12x^Tx$$

এই এনার্জি ল্যান্ডস্কেপে মেমরির আপডেট নিয়মটি হলো: $x^{textnew} = Xcdottextsoftmax(betaX^Tx)$—যা আধুনিক ট্রান্সফরমারের সেলফ-অ্যাটেনশনের হুবহু সমতুল্য! আধুনিক হপফিল্ড নেটওয়ার্কের স্মৃতি ধারণক্ষমতা এক্সপোনেনশিয়ালি বৃদ্ধি পেয়ে দাঁড়ায় $Capprox2^{d/2}$, যা প্রমাণ করে যে ট্রান্সফরমার আসলে একটি আধুনিক অ্যাসোসিয়েটিভ মেমরি এনার্জি সিস্টেম।

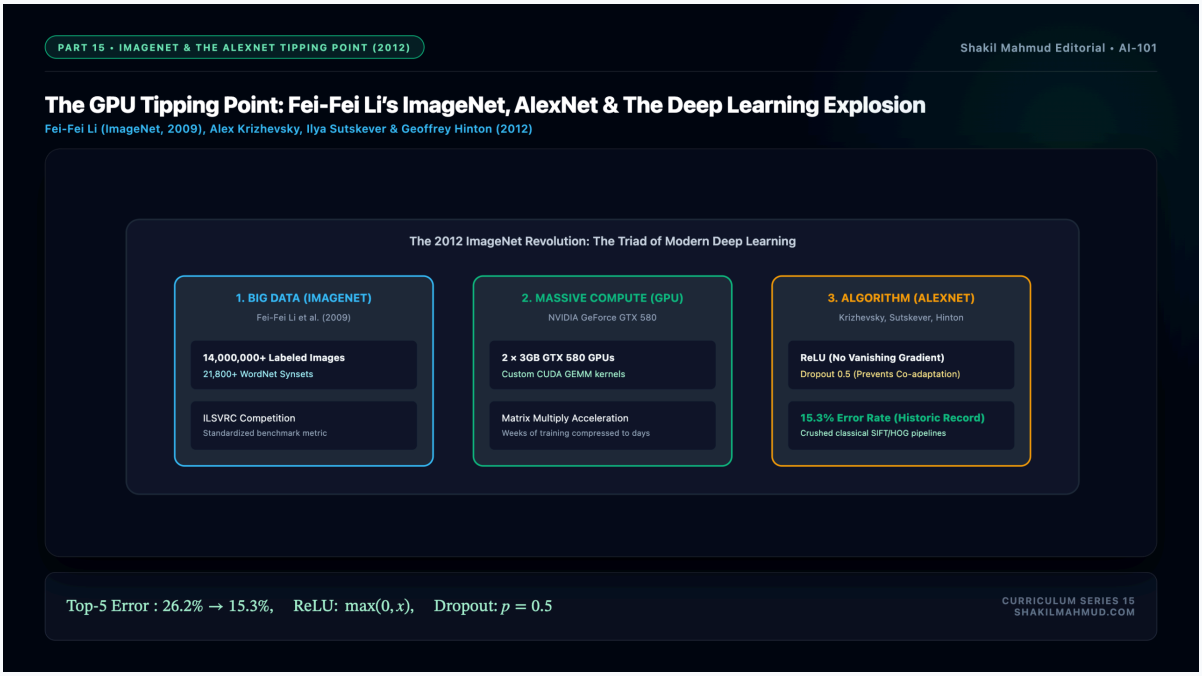
সাইটেশন: মাহমুদ, শাকিল (2026)। "এনার্জি-ভিত্তিক মডেল ও আধুনিক হপফিল্ড নেটওয়ার্ক — বোল্টজম্যান মেশিন থেকে ডেনস মেমরি"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১৪। https://shakilmahmud.com/bn/deep-learning-architectures-part-14-energy-based-models-modern-hopfield-networks-attractor-memory-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • পঞ্চদশ খণ্ড

গ্রাফ নিউরাল নেটওয়ার্ক (GNN) — মেসেজ পাসিং ও ওয়েইসফেইলার-লেহম্যানের সীমা

গ্রাফ নিউরাল নেটওয়ার্ক (GNN) — মেসেজ পাসিং ও ওয়েইসফেইলার-লেহম্যানের সীমা

285 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।*

চিরায়ত ডিপ লার্নিং ইউক্লিডীয় গ্রিডের ওপর প্রতিষ্ঠিত: ছবি হলো পিক্সেলের ২ডি গ্রিড, আর অডিও বা টেক্সট হলো ১ডি ক্রম। কিন্তু প্রাকৃতিক ও সামাজিক জগতের অধিকাংশ উপাত্ত ইউক্লিডীয় নয়: রাসায়নিক অণু, প্রোটিন গঠন, উদ্ধৃতি নেটওয়ার্ক এবং সোশ্যাল মিডিয়া গ্রাফ ইত্যাদি অনিয়মিত টপোলজিক্যাল গ্রাফ $mathcalG = (mathcalV, mathcalE)$ হিসেবে বিদ্যমান। **গ্রাফ নিউরাল নেটওয়ার্ক (GNN)** এই অনিয়মিত স্পেসে ডিপ লার্নিংকে প্রসারিত করে **পারমিউটেশন সমতুল্যতা (Permutation Equivariance)** এবং স্থানীয় **মেসেজ পাসিং (Message Passing)** অ্যালগরিদমের মাধ্যমে।

১. গ্রাফের প্রতিসাম্য ও পারমিউটেশন ইকুইভ্যারিয়েন্স

গ্রাফের নোডগুলোর কোনো নির্দিষ্ট ক্রম বা ইনডেক্স থাকে না। নোডগুলোকে যেকোনো পারমিউটেশন ম্যাট্রিক্স P দিয়ে পুনর্বিন্যস্ত করলেও গ্রাফের অন্তর্নিহিত গঠন অপরিবর্তিত থাকে। ফলে নোড-লেভেল প্রেডিকশনকে P-**Equivariant** হতে হয়: $f(PAP^T, PX) = Pf(A, X)$।

২. মেসেজ পাসিং নিউরাল নেটওয়ার্ক (MPNN)

২০১৭ সালে জাস্টিন গিলমার সমস্ত গ্রাফ আর্কিটেকচারকে একটি সর্বজনীন কাঠামোর মধ্যে আবদ্ধ করেন যা তিনটি ধাপে বিভক্ত:

1. **মেসেজ তৈরি:** প্রতিটি প্রতিবেশী নোড $uinmathcalN(v)$ থেকে মেসেজ m_{uv} তৈরি করা।
2. **মেসেজ একত্রীকরণ (Aggregation):** প্রতিবেশীদের মেসেজগুলোকে প্রতিসম অপারেশনের ($sum, textmean, max$) মাধ্যমে যোগ করা।
3. **নোড আপডেট:** নোডের পূর্ববর্তী অবস্থার সাথে নতুন মেসেজ মিলিয়ে নতুন ভেক্টর h_v নির্ধারণ করা।

Shakil Mahmud • Research Monograph

৩. প্রধান আর্কিটেকচারসমূহ

- **GCN (কিফ ও ওয়েলিং, ২০১৭):** ফার্স্ট-অর্ডার স্পেকট্রাল কনভোলিউশন $H^{(k)} = sigma(tildeD^{-1/2}tildeAtildeD^{-1/2}H^{(k-1)}W)$।
- **GAT (ভেলিকোভিচ, ২০১৮):** প্রতিবেশীদের গুরুত্ব নির্ধারণে গতিশীল সেলফ-অ্যাটেনশন ওজন $alpha_{uv}$ প্রয়োগ।
- **GIN (শু ও সহকর্মীরা, ২০১৯):** মাল্টিসেট পৃথকীকরণে সক্ষম সর্বাধিক শক্তিশালী আর্কিটেকচার।

৪. ওয়েইসফেইলার-লেহম্যান (1-WL) টেস্ট ও প্রকাশ ক্ষমতার সীমা

২০১৯ সালে কেইয়ুলু শু এবং ক্রিস্টোফার মরিস প্রমাণ করেন যে: *যেকোনো মেসেজ পাসিং GNN-এর প্রকাশ ক্ষমতা ক্লাসিক্যাল ১-ওয়েইসফেইলার-লেহম্যান (1-WL) গ্রাফ আইসোমরফিজম টেস্ট দ্বারা সীমাবদ্ধ*। ১-ডাব্লিউএল অ্যালগরিদম যে দুটি ভিন্ন গ্রাফকে আলাদা করতে পারে না (যেমন একটি ৬-নোডের রিং বনাম দুটি ৩-নোডের ত্রিভুজ), কোনো স্ট্যান্ডার্ড GNN-ও তাদের পৃথক করতে পারে না। এই সীমাবদ্ধতা ভাঙতে আধুনিক উচ্চ-ক্রমের সাব-গ্রাফ নেটওয়ার্কের প্রয়োজন হয়।



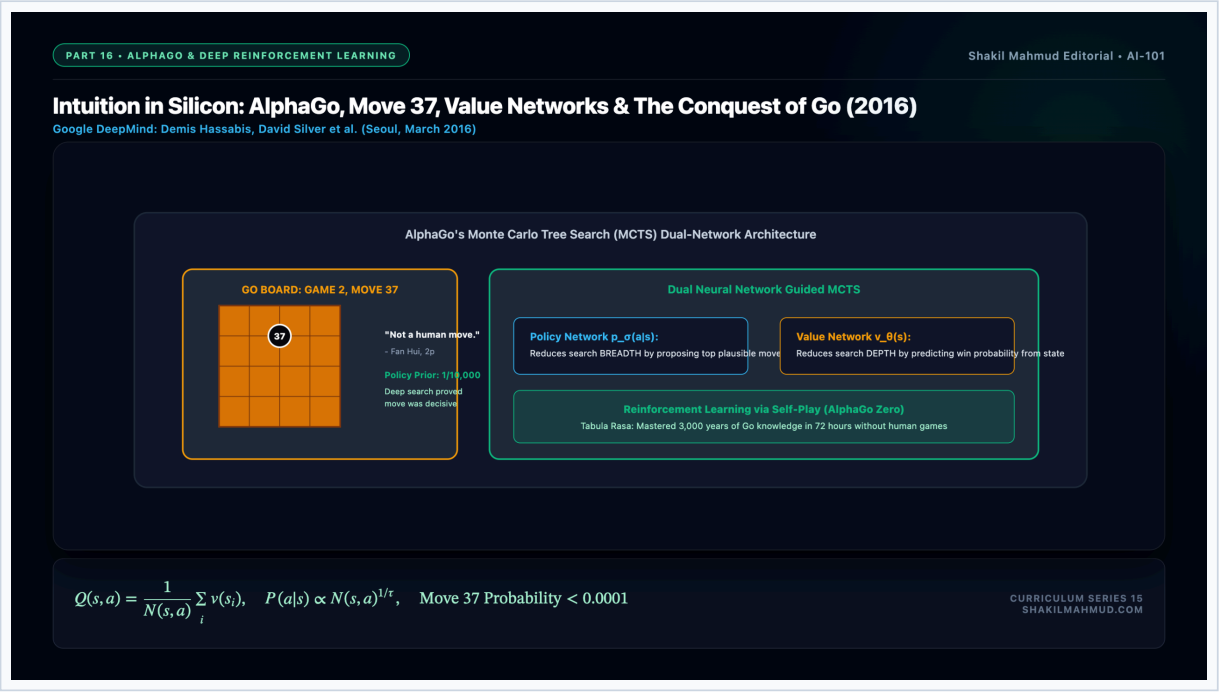
সাইটেশন: মাহমুদ, শাকিল (2026)। "গ্রাফ নিউরাল নেটওয়ার্ক (GNN) — মেসেজ পাসিং ও ওয়েইসফেইলার-লেহম্যানের সীমা"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১৫।
https://shakilmahmud.com/bn/deep-learning-architectures-part-15-graph-neural-networks-message-passing-weisfeiler-lehman-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • ষোড়শ খণ্ড

জ্যামিতিক ডিপ লার্নিং ও স্ফেরিক্যাল সিএনএন — গেজ সিমেট্রি ও ত্রিমাত্রিক ঘূর্ণন সমতুল্যতা

জ্যামিতিক ডিপ লার্নিং ও স্ফেরিক্যাল সিএনএন — গেজ সিমেট্রি ও ত্রিমাত্রিক ঘূর্ণন সমতুল্যতা

285 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।*

২০২১ সালে মাইকেল ব্রনস্টাইন, জোয়ান ব্রুনা, টাকো কোহেন এবং পেটার ভেলিকোভিচ প্রকাশ করেন তাদের ইতিহাসখ্যাত গবেষণা *"Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges"*। ফেলিক্স ক্লাইনের এরলাঙ্গেন প্রোগ্রামের অনুপ্রেরণায় তারা প্রমাণ করেন: সিএনএন, আরএনএন, ট্রান্সফরমার এবং জিএনএন কোনো বিচ্ছিন্ন আবিষ্কার নয়; এগুলো সবই **প্রতিসাম্য গ্রুপ ও জ্যামিতিক অপরিবর্তনীয়তার (Symmetry & Invariance)** সরাসরি গাণিতিক পরিণতি। বাঁকানো রিম্যানিয়ান তল এবং গোলকের ($mathbbS^2$) ওপর **স্ফেরিক্যাল হারমোনিক্স ও গেজ কানেকশন** প্রয়োগের মাধ্যমে জ্যামিতিক ডিপ লার্নিং নিখুঁত ত্রিমাত্রিক ঘূর্ণন সমতুল্যতা ($SO(3)$) প্রতিষ্ঠা করে।

১. জ্যামিতিক ডিপ লার্নিংয়ের পাঁচটি ভিত্তি

যেকোনো নিউরাল আর্কিটেকচার একটি নির্দিষ্ট ডোমেইনের প্রতিসাম্য মেনে চলে:

1. **গ্রিড (Grids):** ইউক্লিডীয় তল ও স্থানান্তর প্রতিসাম্য $implies$ সিএনএন।
2. **গ্রুপ (Groups):** হোমোজেনিয়াস স্পেস $implies$ গ্রুপ-ইকুইভ্যারিয়েন্ট সিএনএন।
3. **গ্রাফ (Graphs):** টপোলজিক্যাল পারমিউটেশন প্রতিসাম্য $implies$ জিএনএন।
4. **জিওডেসিক (Geodesics):** রিম্যানিয়ান ম্যানিফোল্ডের অন্তর্নিহিত দূরত্ব $implies$ ম্যানিফোল্ড নেটওয়ার্ক।
5. **গেজ (Gauges):** ম্যানিফোল্ডের স্থানীয় কোঅর্ডিনেট ফ্রেম রূপান্তর $implies$ গেজ-ইকুইভ্যারিয়েন্ট নেটওয়ার্ক।

Shakil Mahmud • Research Monograph

২. স্ফেরিক্যাল কনভোলিউশন ও ত্রিমাত্রিক ঘূর্ণন গ্রুপ $SO(3)$

জলবায়ু বিজ্ঞান বা মহাকাশবিদ্যায় উপাত্ত থাকে গোলকের উপরিভাগে ($mathbbS^2$)। গোলককে সমতলে মেলে ধরলে মেরু অঞ্চলে বিকৃতি ঘটে। গোলকের স্বাভাবিক ঘূর্ণন প্রতিসাম্য গ্রুপ হলো **স্পেশাল অর্থোগোনাল গ্রুপ** $SO(3)$। টাকো কোহেন উদ্ভাবিত স্ফেরিক্যাল সিএনএন সরাসরি গোলক তলে কনভোলিউশন সম্পন্ন করে ত্রিমাত্রিক ঘূর্ণনের নিখুঁত সমতুল্যতা রক্ষা করে।

৩. হারমোনিক অ্যানালাইসিস ও উইগনার ডি-ম্যাট্রিক্স

গোলকের ওপর কনভোলিউশন গণনায় সাধারণ ফুরিয়ার রূপান্তরের বদলে ব্যবহৃত হয় **স্ফেরিক্যাল হারমোনিক্স (Y_l^m)**। পিটার-ওয়েল উপপাদ্য অনুসারে যেকোনো গোলকীয় ফাংশনকে অর্থোগোনাল হারমোনিক্সে বিভক্ত করা যায়। ঘূর্ণনের অধীনে এই তরঙ্গগুলো *উইগনার ডি-ম্যাট্রিক্স (Wigner D-matrices)* দ্বারা রূপান্তরিত হয়, যা ফ্রিকোয়েন্সি ডোমেনে দ্রুত কনভোলিউশন সম্ভব করে।

৪. গেজ সিমেট্রি ও ফাইবার বান্ডল

জটিল প্রোটিন তলের মতো যেকোনো বাঁকানো ম্যানিফোল্ডে কোনো বৈশ্বিক দিক বা কোঅর্ডিনেট সিস্টেম থাকে না। **গেজ-ইকুইভ্যারিয়েন্ট সিএনএন** স্থানীয় স্পর্শক তলগুলোতে ডিফারেনশিয়াল জ্যামিতির *লেভি-চিভিতা কানেকশন (Levi-Civita connection)* প্রয়োগ করে, যা স্থানীয় কোঅর্ডিনেটের যেকোনো খেয়ালখুশিমতো পরিবর্তনের মুখেও মডেলের আউটপুটকে সম্পূর্ণ অপরিবর্তিত ও নির্ভুল রাখে।

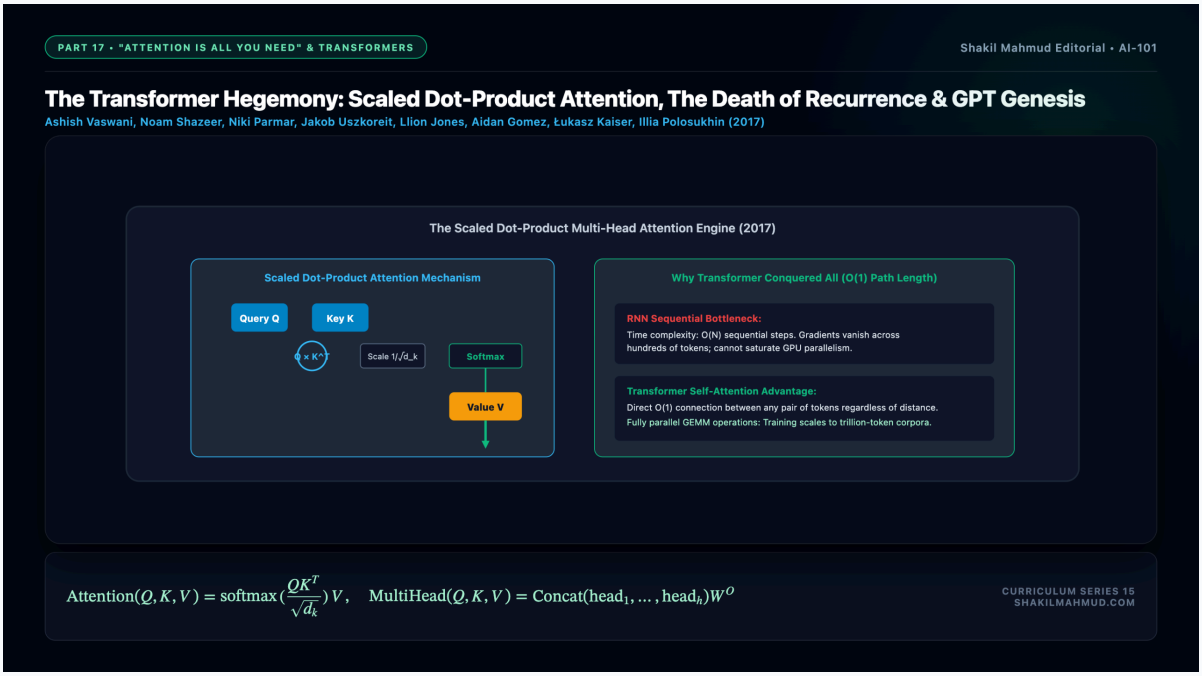
সাইটেশন: মাহমুদ, শাকিল (2026)। "জ্যামিতিক ডিপ লার্নিং ও স্ফেরিক্যাল সিএনএন — গেজ সিমেট্রি ও ত্রিমাত্রিক ঘূর্ণন সমতুল্যতা"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১৬।
https://shakilmahmud.com/bn/deep-learning-architectures-part-16-geometric-deep-learning-spherical-cnns-gauge-symmetries-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • সপ্তদশ খণ্ড

নিউরাল অর্ডিনারি ডিফারেনশিয়াল সমীকরণ — অবিচ্ছিন্ন গভীরতা ও অ্যাডজয়েন্ট পদ্ধতি

নিউরাল অর্ডিনারি ডিফারেনশিয়াল সমীকরণ — অবিচ্ছিন্ন গভীরতা ও অ্যাডজয়েন্ট পদ্ধতি

310 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 রেসিডুয়াল নেটওয়ার্কে তথ্য প্রবাহিত হয় কতগুলো পৃথক, বিচ্ছিন্ন স্তরের মধ্য দিয়ে: $h_{t+1} = h_t + f(h_t)$। ২০১৮ সালে রিকি চেন এবং ডেভিড ডুভেনড উপলব্ধি করেন যে এই রেসিডুয়াল সমীকরণটি আসলে একটি অবিচ্ছিন্ন সাধারণ ব্যবকলনীয় সমীকরণের (ODE) অয়লার ডিসক্রিটাইজেশন মাত্র। সময়ের অতি-ক্ষুদ্র সীমায় ($Deltatto0$) পদার্পণ করে তারা আবিষ্কার করেন **নিউরাল অর্ডিনারি ডিফারেনশিয়াল সমীকরণ (Neural ODEs)**—যা বিচ্ছিন্ন স্তরগুলোকে একটি অবিচ্ছিন্ন ভেক্টর ফিল্ডে রূপান্তরিত করে এবং **অ্যাডজয়েন্ট সেনসিটিভিটি পদ্ধতি** প্রয়োগের মাধ্যমে মেমরি খরচকে $O(1)$ ধ্রুবক মানে নামিয়ে আনে।

১. বিচ্ছিন্ন রেসনেট থেকে অবিচ্ছিন্ন গতিশীল প্রবাহ

রেসিডুয়াল ব্লকের ধাপকে অবিচ্ছিন্ন সময়ের সাপেক্ষে প্রকাশ করলে পাওয়া যায় মূল ডিফারেনশিয়াল সমীকরণ:

$$fracdh(t)dt = f(h(t), t, theta)$$

যেখানে f হলো একটি নিউরাল নেটওয়ার্ক যা স্পেসের প্রতিটি বিন্দুতে বেগ ভেক্টর নির্ধারণ করে। ইনপুট $h(0) = x$ থেকে চূড়ান্ত স্তর T-এ পৌঁছানো আসলে একটি গাণিতিক ইনিশিয়াল ভ্যালু প্রবলেম (IVP):

$$h(T) = h(0) + int_0^T f(h(t), t, theta)dt$$

Shakil Mahmud • Research Monograph

এখানে মডেলের কোনো নির্দিষ্ট “স্তর সংখ্যা” নেই; রানগে-কুট্টা (RK4) বা ডরম্যান্ড-প্রিন্সের মতো আধুনিক ওডিই সলভার ব্যবহারকারীর নির্ধারিত নির্ভুলতার ওপর ভিত্তি করে গতিশীলভাবে ধাপের সংখ্যা নির্ধারণ করে।

২. অ্যাডজয়েন্ট পদ্ধতি ও ধ্রুবক মেমরি ব্যাকপ্রোপাগেশন

সাধারণ ব্যাকপ্রোপাগেশনে ওডিই সলভারের সমস্ত মধ্যবর্তী ধাপ মেমরিতে সংরক্ষণ করতে গেলে জিপিইউ মেমরি দ্রুত ফুরিয়ে যায়। চেন এবং তার দল লেভ পনত্রিয়াগিনের ১৯৬২ সালের ম্যাক্সিমাম নীতি প্রয়োগ করে **অ্যাডজয়েন্ট ডিফারেনশিয়াল সমীকরণ** তৈরি করেন:

$$fracda(t)dt = -a(t)^T fracpartial fpartialh(t), quada(t) = fracpartialLpartialh(t)$$

ব্যাকওয়ার্ড পাসের সময় সলভার সময়কে উল্টোদিকে ($Tto0$) চালিয়ে চূড়ান্ত আউটপুট থেকে ইনপুট পুনর্গঠন করে এবং সাথে সাথে প্যারামিটার গ্রেডিয়েন্ট সংগ্রহ করে। ফলে কোনো মধ্যবর্তী অ্যাক্টিভেশন মেমরিতে জমিয়ে রাখতে হয় না; মডেল যত গভীর বা জটিলই হোক না কেন, ট্রেনিং মেমরি সর্বদা $O(1)$ **ধ্রুবক মানে** স্থির থাকে।

৩. অনিয়মিত সময়ের ডেটায় প্রয়োগ

হাসপাতালের আইসিইউ রেকর্ড বা মহাকাশের সিগন্যালের মতো অনিয়মিত সময়ের ব্যবধানে আসা উপাত্তে সাধারণ আরএনএন বা ট্রান্সফরমার ব্যর্থ হয়। কিন্তু নিউরাল ওডিই অবিচ্ছিন্ন সময়ের মডেল হওয়ায় এটি যেকোনো এলোমেলো সময়ের বিন্দুতে স্বতঃস্ফূর্তভাবে সঠিক পূর্বাভাস দিতে সক্ষম।

সাইটেশন: মাহমুদ, শাকিল (2026)। "নিউরাল অর্ডিন্যারি ডিফারেনশিয়াল সমীকরণ — অবিচ্ছিন্ন গভীরতা ও অ্যাডজয়েন্ট পদ্ধতি"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১৭।
https://shakilmahmud.com/bn/deep-learning-architectures-part-17-neural-odes-continuous-depth-adjoint-state-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • অষ্টাদশ খণ্ড

স্টেট-স্পেস মডেল ও ম্যাম্বা — হিপো প্রজেকশন ও লিনিয়ার-টাইম সিকোয়েন্স

স্টেট-স্পেস মডেল ও ম্যাম্বা — হিপো প্রজেকশন ও লিনিয়ার-টাইম সিকোয়েন্স

310 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।*

আধুনিক লার্জ ল্যাঙ্গুয়েজ মডেলগুলোতে ট্রান্সফরমার রাজত্ব করলেও এর একটি মারাত্মক প্রযুক্তিগত সীমাবদ্ধতা রয়েছে: **কোয়াড্রেটিক বটলনেক ($O(N^2)$)**। সিকোয়েন্সের দৈর্ঘ্য N দ্বিগুণ হলে সেলফ-অ্যাটেনশনের মেমরি ও গণনার খরচ চারগুণ বৃদ্ধি পায়। ফলে লাখ লাখ টোকেনের টেক্সট বা দীর্ঘ ভিডিও প্রসেস করা অত্যন্ত ব্যয়বহুল হয়ে ওঠে। ২০২৩ সালে অ্যালবার্ট গু এবং ট্রি ডাও আবিষ্কার করেন **ম্যাম্বা (Mamba)**। অবিচ্ছিন্ন **স্টেট-স্পেস মডেল (SSM)**-এর ওপর প্রতিষ্ঠিত এই মডেলটি ট্রান্সফরমারের সমমানের বুদ্ধিমত্তা প্রদর্শন করে কিন্তু সময় ও মেমরি খরচকে **লিনিয়ার $O(N)$ জটিলতায়** নামিয়ে আনে।

১. অবিচ্ছিন্ন স্টেট-স্পেস সমীকরণ

একটি অবিচ্ছিন্ন স্টেট-স্পেস মডেল একটি ইনপুট সংকেত $x(t)$-কে একটি লুকায়িত অবস্থা $h(t) in mathbb R^N$-এর মাধ্যমে আউটপুট $y(t)$-এ রূপান্তরিত করে দুটি সাধারণ সমীকরণের সাহায্যে:

$$h'(t) = Ah(t) + Bx(t)$$

$$y(t) = Ch(t) + Dx(t)$$

Shakil Mahmud • Research Monograph

২. হিপো (HiPPO) ফ্রেমওয়ার্ক ও মেমরি সংরক্ষণ

ম্যাট্রিক্স A যদি এলোমেলোভাবে নির্ধারণ করা হয়, তবে মডেল অতীত স্মৃতি দ্রুত ভুলে যায়। ২০২০ সালে অ্যালবার্ট গু আবিষ্কার করেন **HiPPO ফ্রেমওয়ার্ক**। তারা প্রমাণ করেন যে লিজেন্ড্রে অর্থোগোনাল পলিনোমিয়াল (Legendre Polynomials)-এর সহগগুলোকে ট্র্যাক করার মাধ্যমে ম্যাট্রিক্স A এমন একটি রূপ ধারণ করে যা লক্ষাধিক সময়-পদক্ষেপ ধরে অতীতের প্রতিটি সংকেত নিখুঁতভাবে মনে রাখতে পারে।

৩. দ্বৈত রূপ: রিকারেন্স বনাম কনভোলিউশন

স্টেট-স্পেস মডেলকে কম্পিউটারে ব্যবহারের জন্য ডিসক্রিটাইজ করা হলে এর দুটি অসাধারণ দ্বৈত রূপ উন্মোচিত হয়:

- **ইনফারেন্সের সময় রিকারেন্স ($O(1)$):** টোকেন তৈরির সময় এটি একটি সাধারণ আরএনএন-এর মতো কাজ করে, যার ফলে কোনো বিশালাকার KV ক্যাশ মেমরিতে রাখতে হয় না। মেমরি খরচ সম্পূর্ণ ধ্রুবক থাকে।
- **ট্রেনিংয়ের সময় কনভোলিউশন ($O(NlogN)$):** সম্পূর্ণ সিকোয়েন্সকে প্যারালাল ফাস্ট ফুরিয়ার রূপান্তরের (FFT) মাধ্যমে একযোগে প্রশিক্ষণ দেওয়া যায়।

৪. ম্যাম্বা এবং সিলেক্টিভ স্টেট স্পেস (S6)

আগের এসএসএম মডেলগুলো ইনপুট নির্বিশেষে স্থিরভাবে কাজ করত (Time-Invariant)। কিন্তু ম্যাম্বা প্যারামিটারগুলোকে ইনপুটের ওপর নির্ভরশীল করে তোলে: $B_t = f(x_t), C_t = f(x_t), Delta_t = f(x_t)$। এর ফলে অপ্রয়োজনীয় তথ্য ফেলে দিয়ে কেবল গুরুত্বপূর্ণ তথ্য সংরক্ষণ করা সম্ভব হয়। জিপিইউর উচ্চগতির SRAM মেমরিতে **হার্ডওয়্যার-সচেতন প্যারালাল স্ক্যান** বাস্তবায়নের মাধ্যমে ম্যাম্বা লিনিয়ার-টাইমে ফ্ল্যাশ-অ্যাটেনশনের চেয়েও দ্রুত কাজ সম্পন্ন করে।

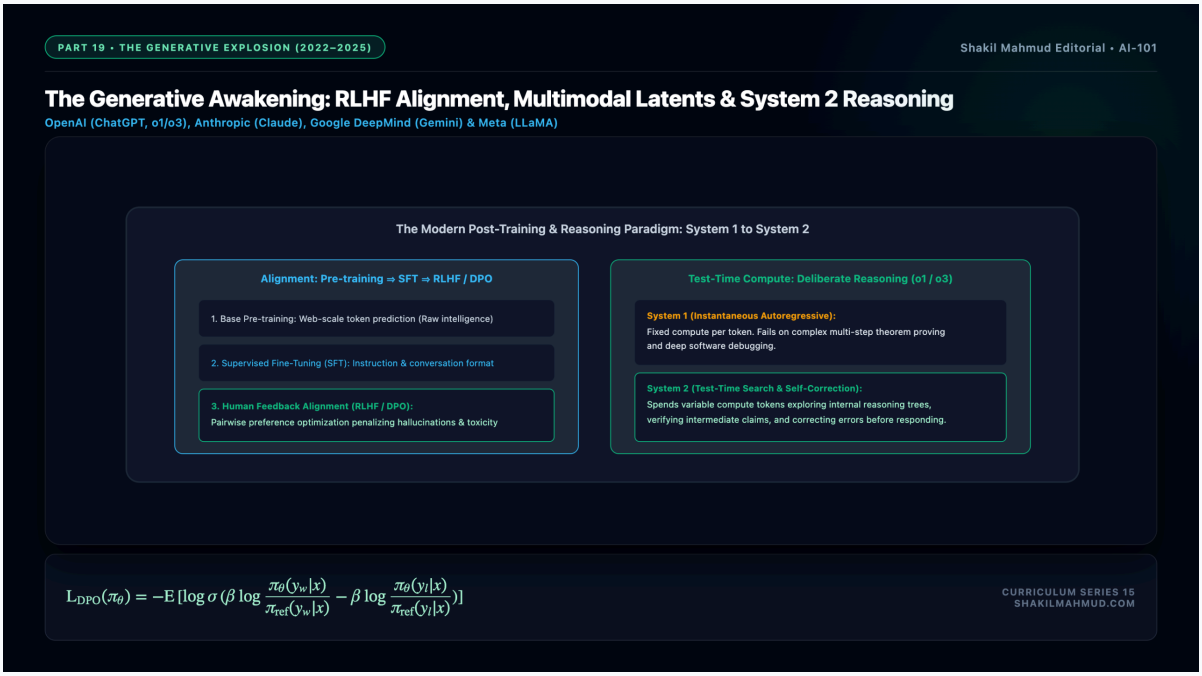
সাইটেশন: মাহমুদ, শাকিল (2026)। "স্টেট-স্পেস মডেল ও ম্যাম্বা — হিপো প্রজেকশন ও লিনিয়ার-টাইম সিকোয়েন্স"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১৮।
https://shakilmahmud.com/bn/deep-learning-architectures-part-18-state-space-models-mamba-linear-time-sequences-bn/

শাকিল মাহমুদ মনোগ্রাফ সিরিজ • উনবিংশ খণ্ড

ডিপ রিইনফোর্সমেন্ট লার্নিং আর্কিটেকচার — ডিপ কিউ-নেটওয়ার্ক থেকে অ্যাক্টর-ক্রিটিক দ্বৈততা

ডিপ রিইনফোর্সমেন্ট লার্নিং আর্কিটেকচার — ডিপ কিউ-নেটওয়ার্ক থেকে অ্যাক্টর-ক্রিটিক দ্বৈততা

304 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 কেবল স্থির ডেটাসেটে কাজ করে যেখানে সঠিক উত্তরের লেবেল আগে থেকেই দেওয়া থাকে। কিন্তু রোবটিক্স, ভিডিও গেমস বা স্বায়ত্তশাসিত সিদ্ধান্ত গ্রহণে কোনো তৈরি উত্তর থাকে না; সেখানে পরিবেশের সাথে ক্রিয়া-প্রতিক্রিয়া ও ভুলের মাধ্যমে শিখতে হয়। **ডিপ রিইনফোর্সমেন্ট লার্নিং (DRL)** বেলম্যানের অপ্টিমালিটি সমীকরণকে ডিপ নিউরাল নেটওয়ার্কের সাথে যুক্ত করে সরাসরি সংবেদনশীল ইনপুট থেকে সর্বোত্তম আচরণিক নীতি তৈরি করে।

১. মার্কভ ডিসিশন প্রসেস ও বেলম্যান সমীকরণ

একটি পরিবেশকে মার্কভ ডিসিশন প্রসেস (MDP) হিসেবে সংজ্ঞায়িত করা হয়। রিচার্ড বেলম্যানের প্রিন্সিপাল অব অপ্টিমালিটি অনুসারে যেকোনো স্টেট-অ্যাকশনের সর্বোত্তম মূল্যমান $Q^*(s, a)$ নির্ধারিত হয়:

Q^*(s, a) = mathcal{R}(s, a) + gamma mathbb{E}left[max_{a’} Q^*(s’, a’) right]

২. রিচার্ড সাটনের “ডেডলি ট্রায়াড” ও ডিপ কিউ-নেটওয়ার্ক (DQN)

ডিপ নিউরাল নেটওয়ার্কের মাধ্যমে কিউ-লার্নিং চালাতে গেলে তিনটি উপাদানের বিপজ্জনক সংমিশ্রণ ঘটে যাকে *ডেডলি ট্রায়াড (Deadly Triad)* বলে—ফাংশন অপ্রক্সিমেশন, বুটস্ট্র্যাপিং এবং অফ-পলিসি লার্নিং। ২০১৫ সালে গুগল ডিপমাইন্ড দুটি যুগান্তকারী স্থাপত্যের মাধ্যমে এই বিপর্যয় প্রতিরোধ করে বিশ্বখ্যাত **DQN** তৈরি করে:

Shakil Mahmud • Research Monograph

- **এক্সপেরিয়েন্স রিপ্লে বাফার (Replay Buffer):** অতীতের সমস্ত অভিজ্ঞতা একটি বাফারে জমা রেখে সেখান থেকে র‍্যান্ডম মিনি-ব্যাচ বেছে নেওয়া হয়, যা সময়ের পারস্পরিক সম্পর্ক ভেঙে প্রশিক্ষণকে স্থিতিশীল করে।
- **টার্গেট নেটওয়ার্ক (Target Network):** লস গণনায় ব্যবহৃত টার্গেট নেটওয়ার্কের প্যারামিটার $theta^-$-কে স্থির রেখে কেবল নির্দিষ্ট সময় পর পর আপডেট করা হয়, যার ফলে চলমান লক্ষ্যের পেছনে ছোটার অস্থিতিশীলতা দূর হয়।

৩. ডাবল ডিকিউএন ও ডুয়েলিং নেটওয়ার্ক

সর্বোচ্চকরণ অপারেশনের কারণে কিউ-ভ্যালুর অতিরিক্ত অনুমানের প্রবণতা দূর করতে **ডাবল ডিকিউএন** সিদ্ধান্ত গ্রহণ ও মূল্যায়নকে আলাদা করে। অন্যদিকে **ডুয়েলিং নেটওয়ার্ক** কিউ-ফাংশনকে স্টেট ভ্যালু $V(s)$ এবং অ্যাডভান্টেজ $A(s,a)$-এ বিভক্ত করে কর্মক্ষমতা বৃদ্ধি করে।

৪. অবিচ্ছিন্ন অ্যাকশন স্পেস ও অ্যাক্টর-ক্রিটিক দ্বৈততা

রোবটের হাত নাড়ানোর মতো অবিচ্ছিন্ন অ্যাকশন স্পেসে $max_a Q(s,a)$ বের করা অসম্ভব। এর সমাধানে আসে **অ্যাক্টর-ক্রিটিক (Actor-Critic)** আর্কিটেকচার: ক্রিটিক নেটওয়ার্ক পরিস্থিতির মান মূল্যায়ন করে, আর অ্যাক্টর নেটওয়ার্ক সরাসরি অ্যাকশন ভেক্টর তৈরি করে। ডেভিড সিলভারের ডিটারমিনিস্টিক পলিসি গ্রেডিয়েন্ট (DPG) উপপাদ্যের ওপর প্রতিষ্ঠিত **DDPG** আধুনিক রোবটিক্স ও কন্ট্রোল সিস্টেমের প্রধানতম ভিত্তি।

সাইটেশন: মাহমুদ, শাকিল (2026)। "ডিপ রিইনফোর্সমেন্ট লার্নিং আর্কিটেকচার — ডিপ কিউ-নেটওয়ার্ক থেকে অ্যাক্টর-ক্রিটিক দ্বৈততা"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ১৯।
https://shakilmahmud.com/bn/deep-learning-architectures-part-19-deep-reinforcement-learning-dqn-actor-critic-architectures-bn/

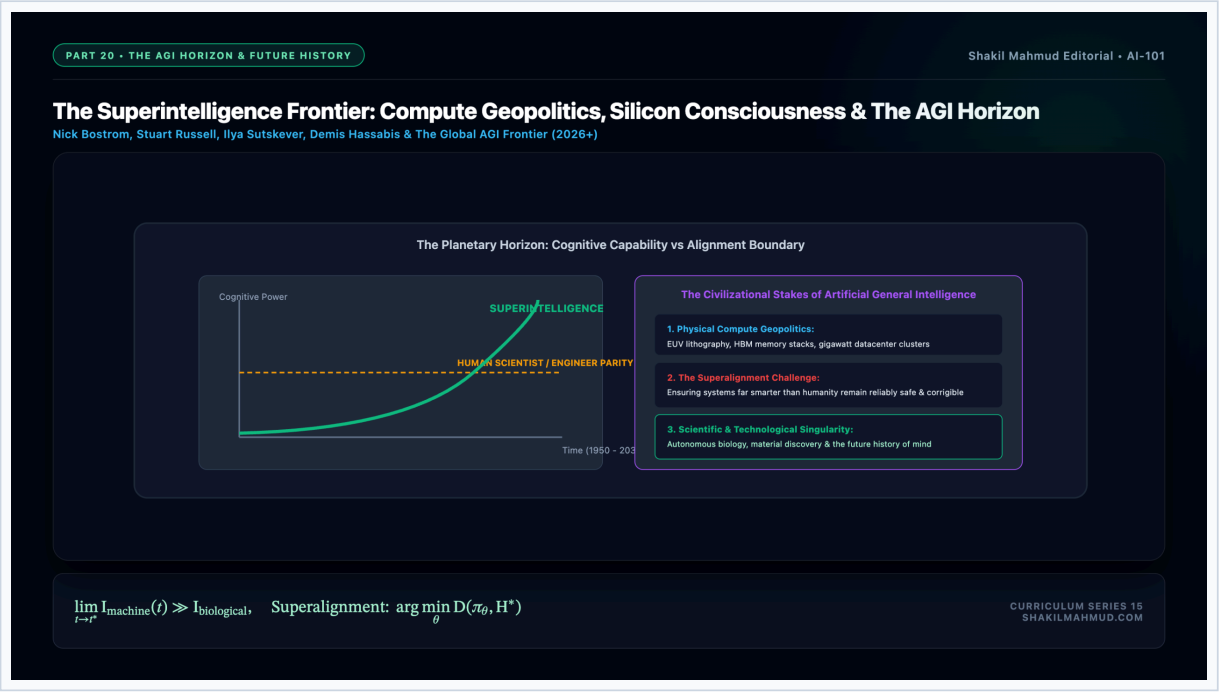
শাকিল মাহমুদ মনোগ্রাফ সিরিজ • বিংশ খণ্ড

ডিপ নেটওয়ার্কের রিপ্রেজেন্টেশন তত্ত্ব — ইনফরমেশন বটলনেক ও নিউরাল কোলাপ্স

ডিপ নেটওয়ার্কের রিপ্রেজেন্টেশন তত্ত্ব — ইনফরমেশন বটলনেক ও নিউরাল কোলাপ্স

341 শব্দ • November 4, 2026 • ওয়েব সংস্করণ

❖



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

কয়েক দশক ধরে গভীর নিউরাল নেটওয়ার্ককে একটি ব্ল্যাক-বক্স হিসেবে দেখা হতো: প্যারামিটার সংখ্যা উপাত্তের চেয়ে বহুগুণ বেশি ($Dggn$) হওয়া সত্ত্বেও এটি কীভাবে কোনো ওভারফিটিং ছাড়া অসাধারণ ভবিষ্যদ্বাণী করে, তা চিরায়ত পরিসংখ্যানিক শিখন তত্ত্বের কাছে ছিল এক বিশাল রহস্য। সাম্প্রতিক বছরগুলোতে নাফতালি তিশবির **ইনফরমেশন বটলনেক নীতি**, ডনোহো ও পাপিয়ানের **নিউরাল কোলাপ্স (Neural Collapse)** এবং **ডাবল ডিসেন্ট** তত্ত্ব ডিপ লার্নিংয়ের এই গভীর গাণিতিক রহস্য উন্মোচন করেছে।

১. ইনফরমেশন বটলনেক নীতি

২০০০ সালে নাফতালি তিশবি প্রস্তাব করেন যে একটি নেটওয়ার্কের মধ্যবর্তী স্তর T একটি ভ্যারিয়েশনাল অপ্টিমাইজেশন সমস্যার সমাধান করে:

min mathcal{L} = I(X; T) – beta I(T; Y)

যেখানে $I(A;B)$ হলো শ্যানন মিউচুয়াল ইনফরমেশন। এই সমীকরণ অনুযায়ী নেটওয়ার্ক দুটি কাজ করে: প্রথমত, ইনপুট X-এর অপ্রয়োজনীয় নয়েজ ও পারিপার্শ্বিক তথ্য মুছে ফেলে ($I(X;T)$ সংকোচন); দ্বিতীয়ত, আউটপুট Y পূর্বাভাসের জন্য প্রয়োজনীয় মূল বৈশিষ্ট্যগুলো অবিকল সংরক্ষণ করে ($I(T;Y)$ সর্বোচ্চকরণ)। প্রশিক্ষণের সময় প্রাথমিক ফিটিং পর্বের পর একটি দীর্ঘ *তথ্য সংকোচন পর্ব (Information Compression Phase)* ঘটে, যা মডেলের স্বাভাবিক সাধারণীকরণ নিশ্চিত করে।

Shakil Mahmud • Research Monograph

২. ডাবল ডিসেন্ট ও ওভারপ্যারামিটারাইজেশন

চিরায়ত পরিসংখ্যানে মডেলের ধারণক্ষমতা বাড়ালে এরর U-আকারে বাড়ে। কিন্তু ২০১৯ সালে মিখাইল বেলকিন প্রমাণ করেন **ডাবল ডিসেন্ট (Double Descent)**: ইন্টারপোলেশন থ্রেশহোল্ড ($D = n$) অতিক্রম করার পর প্যারামিটার সংখ্যা আরও বহুগুণ বাড়ালে টেস্ট এরর পুনরায় মসৃণভাবে হ্রাস পেতে থাকে। অতি-উচ্চ প্যারামিটার স্পেসে গ্রেডিয়েন্ট ডিসেন্ট স্বয়ংক্রিয়ভাবে ন্যূনতম ell_2-নর্মের সমাধান বেছে নেয়।

৩. নিউরাল কোলাপ্স ($mathcalNC$): প্রশিক্ষণের চূড়ান্ত জ্যামিতি

২০২০ সালে ভার্দান পাপিয়ান এবং ডেভিড ডনোহো আবিষ্কার করেন যে ট্রেনিং এরর শূন্য হওয়ার পর অতিরিক্ত প্রশিক্ষণ চালিয়ে গেলে শেষ স্তরের ফিচার এবং ক্লাসিফায়ার ওয়েটগুলো একটি সর্বজনীন জ্যামিতিক রূপ ধারণ করে যাকে বলে **নিউরাল কোলাপ্স**:

1. **ভেরিয়্যাবিলিটি কোলাপ্স ($mathcalNC_1$):** একই ক্লাসের সমস্ত ডেটা পয়েন্ট তাদের নিজস্ব ক্লাস সেন্ট্রয়েডে সংকুচিত হয়ে যায় ($Sigma_W tomathbf0$)।
2. **সিমপ্লেক্স ইটিএফ ($mathcalNC_2$):** বিভিন্ন ক্লাসের সেন্ট্রয়েডগুলো একটি *Simplex Equiangular Tight Frame* তৈরি করে—যেখানে প্রতিটি ক্লাসের মধ্যবর্তী কোণ সর্বোচ্চ এবং হুবহু সমান (যেমন ৩টি ক্লাসের জন্য ১২০° সমবাহু ত্রিভুজ, ৪টি ক্লাসের জন্য ১০৯.৫° টেট্রাহেড্রন)।
3. **স্ব-দ্বৈততা ও অ্যালাইনমেন্ট ($mathcalNC_3$):** ক্লাসিফায়ারের ওজন ভেক্টরগুলো ক্লাস সেন্ট্রয়েডের সাথে নিখুঁতভাবে সারিবদ্ধ হয়।

নিউরাল কোলাপ্স প্রমাণ করে যে ডিপ লার্নিংয়ের চরম সীমায় সমস্ত বৈচিত্র্য বিলুপ্ত হয়ে এক অনন্য প্রতিসম জ্যামিতিক কাঠামোয় আত্মপ্রকাশ করে।

সাইটেশন: মাহমুদ, শাকিল (2026)। "ডিপ নেটওয়ার্কের রিপ্রেজেন্টেশন তত্ত্ব — ইনফরমেশন বটলনেক ও নিউরাল কোলাপ্স"। ডিপ লার্নিং ও রিপ্রেজেন্টেশন তত্ত্ব, খণ্ড ২০।
https://shakilmahmud.com/bn/deep-learning-architectures-part-20-representation-theory-information-bottleneck-neural-collapse-bn/

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:

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning*. Nature, 521(7553): 436–444.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. CVPR 2016: 770–778.

Kingma, D. P., & Welling, M. (2013). *Auto-Encoding Variational Bayes*. arXiv:1312.6114 [stat.ML].

Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*. arXiv:2104.13478 [cs.LG].

Gu, A., & Dao, T. (2023). *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*. arXiv:2312.00752 [cs.LG].

Shakil Mahmud • Research Monograph

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