

ভূমিকা: মেশিন লার্নিং মডেলের প্রকারভেদ

পূর্বে মেশিন লার্নিং-এর প্রকারভেদ আলোচনা করা হয়েছিল তত্ত্বাবধানের পরিমাণের (Amount of Supervision) উপর ভিত্তি করে (যেমন: সুপারভাইজড, আনসুপারভাইজড, সেমি-সুপারভাইজড, রিইলফোর্মেন্ট লার্নিং)। কিন্তু আজ যে প্রকারভেদ আলোচনা করা হচ্ছে, তা নির্ভর করে মেশিন লার্নিং মডেলটি কীভাবে প্রশিক্ষিত হয় এবং উৎপাদন পরিবেশে (**Production Environment**) কীভাবে কাজ করে।

প্রোডাকশন এনভায়রনমেন্ট কী?

সহজ ভাষায়, উৎপাদন (**Production**) হলো সেই সার্ভার, যেখানে আপনার তৈরি করা সফটওয়্যার কোডটি চলমান থাকে। যখন আপনি একটি মেশিন লার্নিং সফটওয়্যার তৈরি করেন, তখন সেটি আপনার মেশিনে না রেখে একটি সার্ভারে স্থাপন (deploy) করা হয়। ব্যবহারকারীরা সেই সার্ভারের আইপি অ্যাড্রেস ব্যবহার করে তাদের প্রশ্নের উত্তর পেতে পারে। এই চলমান সার্ভারের পরিবেশকেই উৎপাদন পরিবেশ (**Production Environment**) বলা হয়।

মেশিন লার্নিং সার্ভারে কীভাবে আচরণ করে, তার ভিত্তিতে দুটি ভিন্ন প্রকারের লার্নিং প্রক্রিয়া নিয়ে আলোচনা করা হয়েছে:

- ব্যাচ মেশিন লার্নিং (**Batch Machine Learning**) বা অফলাইন লার্নিং (**Offline Learning**)।
- অনলাইন মেশিন লার্নিং (**Online Machine Learning**) বা অনলাইন লার্নিং (**Online Learning**)।

১. ব্যাচ লার্নিং বা অফলাইন লার্নিং (**Batch Learning or Offline Learning**)

ব্যাচ লার্নিং হলো মেশিন লার্নিং মডেলকে প্রশিক্ষণ দেওয়ার একটি প্রথাগত বা প্রচলিত (**conventional**) পদ্ধতি।

ব্যাচ লার্নিং-এর প্রশিক্ষণ প্রক্রিয়া

- পুরো ডেটার ব্যবহার: ব্যাচ লার্নিং-এর ফেরে, মডেলকে প্রশিক্ষণ দেওয়ার জন্য পুরো ডেটাসেটটি একসাথেই ব্যবহার করা হয়।
- অ-বৃদ্ধিভিত্তিক প্রশিক্ষণ (**No Incremental Training**): এখানে ডেটা ছোট ছোট টুকরোগুলিতে ভাগ করে বা বৃদ্ধিভিত্তিক ভাবে প্রশিক্ষণ দেওয়া হয় না। মডেলকে পুরো ডেটার লট (batch) নিয়ে প্রশিক্ষণ করানো হয়।
- অফলাইন প্রশিক্ষণ: সাধারণত, যদি আপনি বাস্তব বিশ্বের সমস্যা নিয়ে কাজ করেন এবং ডেটার পরিমাণ বড় হয়, তবে এত বড় ডেটা সার্ভারে লোড করে প্রশিক্ষণ দেওয়া ব্যবহৃত (**costly**) এবং সময়সাপেক্ষ (**time-taking**) হয়। তাই ডেটা বিজ্ঞানী বা ইঞ্জিনিয়াররা এই প্রশিক্ষণ অফলাইনে, সাধারণত তাদের নিজস্ব মেশিনে সম্পন্ন করেন।
- স্থাপন (**Deployment**): একবার মডেল প্রশিক্ষণ হয়ে গেলে এবং সেটির কার্যকারিতা পরীক্ষা (test) করা হলে, শেখা মডেলটিকে (**trained model**) সার্ভারে নিয়ে গিয়ে স্থাপন (deploy) করা হয়।

উদাহরণস্বরূপ, নেটক্লিওর জন্য একটি সুপারিশ ইঞ্জিন (recommendation engine) যদি ব্যাচ লার্নিং ব্যবহার করে তৈরি করা হয়, তবে সেটিকে প্রথমে নিজের মেশিনে সম্পূর্ণভাবে প্রশিক্ষণ দেওয়া হয় এবং তারপর নেটক্লিওর সার্ভারে স্থাপন করা হয়। সেখান থেকেই মডেলটি ব্যবহারকারীদের মুভি সুপারিশ করতে থাকে।

ব্যাচ লার্নিং-এর প্রধান সীমাবদ্ধতা: স্থির মডেল (**The Static Model**)

ব্যাচ লার্নিং-এর মাধ্যমে প্রশিক্ষিত মডেলের একটি প্রধান সমস্যা হলো, একবার শেখার পর মডেলটি স্থির (**static**) হয়ে যায়। এটি পুরোনো তথ্যের ভিত্তিতে কাজ করতে থাকে। কিন্তু সময়ের সাথে সাথে ব্যবসার পরিস্থিতি (**business scenario**) এবং তথ্য ক্রমাগত বিকশিত হতে থাকে।

যদি একটি স্প্যাম ক্লাসিফায়ার আজকের ডেটা নিয়ে তৈরি হয়, কিন্তু পরবর্তী এক বছরে স্প্যামাররা নতুন কৌশল অবলম্বন করে, তবে পুরোনো মডেলটি সেই নতুন স্প্যাম শনাক্ত করতে ব্যর্থ হবে এবং অকেজে (obsolete) হয়ে যাবে। মডেলটিকে কার্যকর রাখার জন্য নতুন ডেটা অনুযায়ী ক্রমাগত আপডেট হতে থাকা প্রয়োজন।

মডেল পুনরায় প্রশিক্ষণের চক্র (**The Re-training Cycle**)

মডেলটিকে প্রাসঙ্গিক এবং আপ-টু-ডেট রাখার জন্য ডেটা বিজ্ঞালীরা পর্যায়ক্রমে মডেলকে পুনরায় প্রশিক্ষণ (**re-train**) করান। এই প্রক্রিয়াটি একটি চক্রাকারে পুনরাবৃত্তি হয়:

1. নতুন ডেটা সংগ্রহ করা হয় (যা ক্রমাগত আপডেট হতে থাকে)।
2. সংগৃহীত নতুন ডেটা এবং পুরোনো ডেটাকে একসাথে মার্জ (merge) বা একত্রিত করা হয়।
3. মডেলটিকে আবার প্রথম থেকে সম্পূর্ণভাবে প্রশিক্ষণ দেওয়া হয়।
4. প্রশিক্ষণের পর মডেলটি পুনরায় পরীক্ষা করা হয়।
5. সবশেষে, আপডেট করা মডেলটি পুনরায় সার্ভারে স্থাপন করা হয়।

এই পুরো চক্রটি একটি নির্দিষ্ট সময়সীমা পর পর (যেমন ২৪ ঘণ্টা পর, প্রতি সপ্তাহে, বা প্রতি মাসে, এমনকি ছয় মাস অন্তর) পুনরাবৃত্তি হতে থাকে। যদিও লোকেরা দ্রুত আপডেট করতে চায়, ডেটার পরিমাণের উপর নির্ভর করে এই সময়সীমা নির্ধারিত হয়।

২. ব্যাচ লার্নিং-এর অসুবিধা ও যেখানে এটি ব্যর্থ হতে পারে (**Disadvantages and Limitations**)

ব্যাচ লার্নিং বেশিরভাগ ক্ষেত্রেই ব্যবহৃত হলেও, কিছু পরিস্থিতিতে এটি সমস্যার সৃষ্টি করে বা অকার্যকর প্রমাণিত হয়:

ক. ডেটার আকার সংক্রান্ত সীমাবদ্ধতা (**Data Size Limitation**)

যদি ডেটার পরিমাণ অত্যন্ত দ্রুত হারে বাঢ়তে থাকে এবং বিপুল পরিমাণে জমে যায় (যেমন সোশ্যাল নেটওয়ার্কিং ডেটা), তবে এটি একটি সীমাবদ্ধতা সৃষ্টি করে। যেহেতু ব্যাচ লার্নিং-এ পুরো ডেটার স্থূপটিকে (full chunk) একবারে নিয়ে প্রশিক্ষণ দিতে হয়, ডেটা অতিরিক্ত বড় হলে প্রশিক্ষণ সরঞ্জামগুলো (training tools) পুরো ডেটা প্রক্রিয়াকরণ (process) করতে নাও পারতে পারে।

খ. হার্ডওয়্যার ও সংযোগের সীমাবদ্ধতা (**Hardware and Connectivity Limitation**)

যদি আপনার মেশিন লার্নিং মডেলটি এমন পরিবেশে চলে যেখানে সার্ভারের সাথে ঘন ঘন এবং তাৎক্ষণিক সংযোগ (instant connectivity) স্থাপন করা সম্ভব নয়, তবে মডেল আপডেট করা অসম্ভব হয়ে পড়ে।

- **উদাহরণ:** যদি সামরিক বাহিনীর ব্যবহারের জন্য চরম পরিস্থিতিতে (যেমন ল্যাদাখের মতো জায়গায়) কোনো মেশিন লার্নিং অ্যাপ তৈরি করা হয়, অথবা কোনো স্যাটেলাইটে স্থাপন করা হয়, যেখানে ইন্টারনেট বা সংযোগের অভাব রয়েছে। একবার মডেল স্থাপন হওয়ার পর, তাকে ফেরত আনা, নতুন ডেটা দিয়ে প্রশিক্ষণ দেওয়া এবং

পুনরায় আপডেট করা কঠিন বা অসম্ভব হয়ে যায়, যতক্ষণ না তারা ইল্টারনেট সংযোগযুক্ত পরিবেশে ফিরে আসে। এই ধরনের পরিস্থিতিতে ব্যাচ লার্নিং ব্যর্থ হতে পারে।

গ. তাংক্ষণিক প্রতিক্রিয়ার অভাব বা বিলম্ব (**Latency and Stale Response**)

ব্যাচ লার্নিং-এর মডেলগুলি সাধারণত ২৪ ঘন্টা বা তার চেয়ে বেশি সময় পর আপডেট হয়।

- যদি আপনি একটি সোশ্যাল নেটওয়ার্কিং ওয়েবসাইটে ট্রেন্ডিং স্টোরি দেখানোর জন্য মডেল তৈরি করেন এবং সেটি ২৪ ঘন্টা পর পর আপডেট হয়।
- ধরুন হঠাৎ করে নেটওয়ার্কে ডিমনিটাইজেশন (Demonetization) নিয়ে কোনো গুরুত্বপূর্ণ বা দ্রুত ছড়িয়ে পড়া খবর এলো। ব্যবহারকারীরা তাংক্ষণিকভাবে সেই বিষয়ে আগ্রহী হয়ে উঠলেও, যেহেতু সিস্টেমটি ব্যাচ লার্নিং ব্যবহার করছে, তাই সেটি ২৪ ঘন্টা পর্যন্ত সেই নতুন আগ্রহ বা ডেটা প্রক্রিয়া করতে পারে না।
- ফলস্বরূপ, ব্যবহারকারীর আগ্রহের আপডেটটি ২৪ ঘন্টা পরে ঘটবে। ততক্ষণে খবরটি হয়তো তার গুরুত্ব হারিয়েছে, এবং সিস্টেম আপডেট হওয়ার পর ব্যবহারকারীর ফিড অপযোজনীয় বা পুরনো ডিমনিটাইজেশনের খবরে ভরে যেতে পারে। এই দীর্ঘ বিলম্বের কারণে, ব্যাচ লার্নিং অনেক ক্ষেত্রে তাংক্ষণিক প্রয়োজনীয় তথ্য সরবরাহ করতে ব্যর্থ হয়।

এই অসুবিধাগুলির কারণে, এমন কিছু পরিস্থিতি আছে যেখানে ব্যাচ লার্নিং সহায়ক নয় এবং সমস্যা তৈরি করতে পারে। এই ধরনের পরিস্থিতি মোকাবেলার জন্যই অনলাইন লার্নিং পদ্ধতির প্রয়োজন হয়।

<https://vowpalwabbit.org/>

<https://riverml.xyz/dev/>