

১. ভূমিকা ও প্রেক্ষাপট (Introduction and Context)

মেশিন লার্নিং-এর এমন প্রকারভেদ নির্ভর করে একটি মেশিন লার্নিং মডেল কীভাবে প্রশিক্ষিত হয় এবং কীভাবে উৎপাদন পরিবেশে (Production Environment) কাজ করে।

উৎপাদন পরিবেশ কী? (**What is Production Environment?**) সরল ভাষায়, উৎপাদন বা প্রোডাকশন হলো সেই সার্ভার যেখানে আপনার সফটওয়্যার কোডটি চলছে। যখন আপনি কোনো মেশিন লার্নিং সফটওয়্যার তৈরি করেন এবং সেটিকে সার্ভারে স্থাপন (deploy) করেন, তখন সেই সার্ভারের চলমান পরিবেশকে উৎপাদন পরিবেশ বলা হয়। ব্যবহারকারীরা সেই সার্ভারের আইপি অ্যাড্রেস ব্যবহার করে তাদের প্রশ্নের উত্তর পেতে পারে।

মেশিন লার্নিং মডেল সার্ভারে কীভাবে কাজ করে বা আচরণ করে, তার ভিত্তিতে সোর্সগুলোতে দুটি প্রধান ধরনের মডেল প্রশিক্ষণের কথা বলা হয়েছে:

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

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

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

প্রশিক্ষণের প্রক্রিয়া (**Training Process**)

- পুরো ডেটার ব্যবহার: ব্যাচ লার্নিং-এর ক্ষেত্রে, মডেলকে প্রশিক্ষণ দেওয়ার জন্য পুরো ডেটাসেটটি একসাথে ব্যবহার করা হয়।
- অ-বৃদ্ধিভিত্তিক প্রশিক্ষণ (**No Incremental Training**): এখানে ডেটা ছেট ছেট অংশে বিভক্ত করে বা বৃদ্ধিভিত্তিক ভাবে প্রশিক্ষণ দেওয়া হয় না। বরং, পুরো ডেটার লট (batch) নিয়েই প্রশিক্ষণ করানো হয়।
- অফলাইন প্রশিক্ষণ: সাধারণত, যখন ডেটার পরিমাণ খুব বড় হয়, তখন পুরো ডেটা সার্ভারে লোড করে প্রশিক্ষণ দেওয়া অত্যন্ত ব্যয়বহুল (costly) এবং সময়সাপেক্ষ (time-taking) হয়। তাই ডেটা বিজ্ঞানী বা ইঞ্জিনিয়াররা এই প্রশিক্ষণ অফলাইনে (সাধারণত তাদের নিজেদের মেশিনে) করে থাকেন।
- স্থাপনা (**Deployment**): একবার মডেল প্রশিক্ষণ হয়ে গেলে এবং সেটির কার্যকারিতা পরীক্ষা (test) করা হলে, শেখা মডেলটিকে সার্ভারে নিয়ে গিয়ে স্থাপন (deploy) করা হয়।
 - উদাহরণ: আপনি আপনার মেশিনে নেটক্লিয়ের জন্য একটি সুপারিশ ইঞ্জিন (recommendation engine) তৈরি ও প্রশিক্ষণ দিলেন, এবং প্রশিক্ষণ শেষে সেটিকে নেটক্লিয়ের সার্ভারে রাখলেন, যাতে ব্যবহারকারীরা এলে সিস্টেম স্বয়ংক্রিয়ভাবে তাদের সিনেমা সুপারিশ করতে পারে।

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

ব্যাচ লার্নিং-এর একটি বড় সমস্যা হলো, একবার শেখার পর মডেলটি স্থির (**static**) হয়ে যায়। যদিও এটি পুরনো তথ্যের ভিত্তিতে কাজ করতে থাকে, কিন্তু ব্যবসার পরিস্থিতি এবং তথ্য ক্রমাগত বিকশিত (evolve) হতে থাকে।

- যদি কোনো স্প্যাম ক্লাসিফায়ার আজকের ডেটা নিয়ে তৈরি হয়, কিন্তু ভবিষ্যতে স্প্যামাররা নতুন কৌশল (new techniques) অবলম্বন করে, তবে পুরোনো মডেলটি অকেজো (obsolete) হয়ে যেতে পারে।
- মডেলটিকে প্রাসঙ্গিক রাখতে, নতুন ডেটা অনুযায়ী মডেলের ক্রমাগত আপডেট হতে থাকা প্রয়োজন।

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

যেহেতু ব্যাচ লার্নিং বৃক্ষিভিত্তিক হয় না, তাই মডেলকে আপডেট করার জন্য পর্যবেক্ষণে বারবার পুনরায় প্রশিক্ষণ (re-train) করাতে হয়।

- নতুন ডেটা সংগ্রহ করা হয়।
- নতুন ডেটা এবং পুরো ডেটা একত্রিত (merge) করা হয়।
- মডেলটিকে আবার প্রথম থেকে সম্পূর্ণ প্রশিক্ষণ দেওয়া হয়।
- মডেলটি পুনরায় পরীক্ষা করা হয়।
- সব শেষে, আপডেট করা মডেলটি পুনরায় সার্ভারে স্থাপন করা হয়।

এই পুরো প্রক্রিয়াটি একটি চক্রাকারে একটি নির্দিষ্ট সময়সীমা পর পর (যেমন ২৪ ঘণ্টা, এক সপ্তাহ, এক মাস বা ছয় মাস) পুনরাবৃত্তি হতে থাকে।

৩. ব্যাচ লার্নিং-এর অসুবিধা ও সীমাবদ্ধতা (Disadvantages of Batch Learning)

ব্যাচ লার্নিং সাধারণত ব্যবহৃত হলেও, কিছু পরিস্থিতিতে এটি সমস্যার সৃষ্টি করে বা সম্পূর্ণভাবে ব্যর্থ হতে পারে:

ক. ডেটার আকার সংক্রান্ত সীমাবদ্ধতা (Data Size Limitation)

যদি ডেটার পরিমাণ অত্যন্ত দ্রুত হারে বাড়তে থাকে এবং বিপুল পরিমাণে জমে যায় (যেমন সোশ্যাল নেটওয়ার্কিং ডেটা), তবে আপনার প্রশিক্ষণ সরঞ্জামগুলো (training tools) পুরো ডেটার স্লুপটিকে (chunk) একবারে প্রক্রিয়াকরণ (process) করতে নাও পারতে পারে। যেহেতু ব্যাচ লার্নিং-এ পুরো ডেটা নিয়ে প্রশিক্ষণ দিতে হয়, তাই ডেটা অতিরিক্ত বড় হলে এটি কাজ করা কঠিন হয়ে পড়ে।

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

যদি আপনার মডেলটি এমন পরিবেশে চল যেখানে সার্ভারের সাথে ঘন ঘন সংযোগ (frequent connectivity) স্থাপন করা সম্ভব নয়, তবে আপডেট করা অসম্ভব হয়ে পড়ে।

- উদাহরণ: যদি সামরিক বাহিনীর ব্যবহারের জন্য চরম পরিস্থিতিতে (যেমন ল্যাদাখের মতো জায়গায়) কোনো মেশিন লার্নিং অ্যাপ তৈরি করা হয় বা কোনো স্যাটেলাইটে স্থাপন করা হয়, যেখানে ইন্টারনেট সংযোগের অভাব রয়েছে, তবে মডেলটিকে সহজে ফিরিয়ে আনা, নতুন ডেটা দিয়ে প্রশিক্ষণ দেওয়া এবং পুনরায় স্থাপন করা যায় না। এমন ক্ষেত্রে ব্যাচ লার্নিং ব্যর্থ হয়।

গ. তাঁক্ষণিক প্রতিক্রিয়ার অভাব (Stale Response/Latency)

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

- যদি হঠাতে করে নেটওয়ার্কে কোনো গুরুত্বপূর্ণ বা দ্রুত ছাড়িয়ে পড়া থবর (যেমন ডিমলিটাইজেশনের থবর) আসে, তবে আপনার মডেল (যেমন ট্রেন্ডিং স্টেটির দেখানোর মডেল) ২৪ ঘণ্টা ধরে সেই বিষয়ে ব্যবহারকারীর আগ্রহ সম্পর্কে জানতে পারে না বা তাদের ব্যক্তিগত ফিড আপডেট করতে পারে না।
- এই দীর্ঘ বিলম্বের কারণে, যখন ২৪ ঘণ্টা পরে সিস্টেম আপডেট হয়, তখন হয়তো থবরটি প্রাসঙ্গিকভা হারিয়েছে বা ব্যবহারকারীর ফিডটি অতিরিক্ত অপ্রয়োজনীয় তথ্য ভরে গেছে। এই পরিস্থিতিতে ব্যাচ লার্নিং ব্যবহারকারীদের সঠিক সময়ে প্রয়োজনীয় তথ্য দিতে পারে না।

এই সমস্যাগুলো এডানোর জন্য এবং দ্রুত পরিবর্তিত পরিস্থিতিতে মডেলকে কার্যকর রাখার জন্য, এমন কিছু পরিস্থিতি আছে যেখানে ব্যাচ লার্নিং সহায়ক নয়। এই ধরনের পরিস্থিতিতেই অনলাইন লার্নিং (**Online Learning**) নামে অন্য একটি পদ্ধতি ব্যবহার করা হয়।