

এই আলোচনার ভিত্তি হলো মডেলটি কীভাবে জ্ঞান অর্জন করে বা শেখে। এটি একটি গুরুত্বপূর্ণ প্রকারভেদ, যা আমাদের বুঝতে সাহায্য করে যে একটি অ্যালগরিদম ডেটার সাথে কীভাবে মিথস্ক্রিয়া করে।

মেশিন লার্নিং মডেলের শেখার প্রক্রিয়া (How ML Models Learn)

মেশিন লার্নিং মডেলগুলো প্রধানত দুটি উপায়ে শেখে, যা অনেকটা মানুষের শেখার প্রক্রিয়ার মতো:

- মুখ্য করা (Memorizing): শুধু তথ্যগুলো মনে রাখা বা মুখ্য করে নেওয়া।
- সাধারণীকরণ বা ধারণা বোঝা (Generalizing): একটি অন্তর্নিহিত নীতি বা ধারণা (underlying principle) বোঝার চেষ্টা করা।

এই দুটি পদ্ধতির উপর ভিত্তি করে মেশিন লার্নিং অ্যালগরিদমগুলোকে দুটি প্রধান ভাগে ভাগ করা হয়:

- ইনস্ট্যান্স-বেজড লার্নিং (Instance-Based Learning)
- মডেল-বেজড লার্নিং (Model-Based Learning)

১. ইনস্ট্যান্স-বেজড লার্নিং (Instance-Based Learning)

ইনস্ট্যান্স-বেজড লার্নিং হলো সেই পদ্ধতি, যেখানে অ্যালগরিদম ডেটা পয়েন্টগুলোকে আক্ষরিক অর্থে ধরে রাখে বা মুখ্য করে।

কাজের প্রক্রিয়া ও উদাহরণ

ধরুন আপনার কাছে কিছু ডেটা আছে, যেমন: ছাত্রছাত্রীদের আইকিউ, সিজিপিএ এবং তাদের প্লেসমেন্ট হয়েছে কি হয়নি (হ্যাঁ বা না)। এটি একটি শ্রেণিবিন্যাস বা ক্লাসিফিকেশন সমস্যা।

ইনস্ট্যান্স-বেজড অ্যালগরিদম এই সমস্যাটিকে এভাবে মোকাবিলা করে:

- ডেটা সংরক্ষণ: এটি প্রশিক্ষণ ডেটাগুলোকে (Training Data) যেমন আছে, ভবত্ত নিজের কাছে সংরক্ষণ করে।
- স্কোরিং বা ভবিষ্যত্বাণী: যখন কোনো নতুন (incoming) ছাত্রের ডেটা আসে এবং তার প্লেসমেন্ট হবে কিনা জানতে চাওয়া হয় (এই নতুন পয়েন্টকে স্কোরিং ইনস্ট্যান্স বলা হয়), তখন অ্যালগরিদমটি কোনো মডেল বা রূল ব্যবহার করে না।
- সাদৃশ্য নির্ভরতা: এটি কেবল সাদৃশ্য (Similarity) বা দূরত্ব (Distance)-এর উপর নির্ভর করে। অ্যালগরিদমটি নতুন পয়েন্টটির সাথে ডেটাসেটের অন্য সব পয়েন্টের দূরত্ব বা সাদৃশ্য বের করে।
- প্রতিবেশী সিদ্ধান্ত: এটি নতুন পয়েন্টের নিকটতম কিছু সংখ্যক প্রতিবেশী (Nearest Neighbors) পয়েন্ট ধরে নেয় (যেমন, ৩ জন বা ৫ জন প্রতিবেশী)। যদি এই নিকটতম প্রতিবেশীদের মধ্যে বেশিরভাগেরই প্লেসমেন্ট হয়ে থাকে, তবে অ্যালগরিদম সিদ্ধান্ত নেয় যে নতুন ছাত্রটিরও প্লেসমেন্ট হবে।

মূল ধারণা: ইনস্ট্যান্স-বেজড লার্নিং-এর মূল নীতি হলো: আপনি যেখানে থাকেন, সেটি আপনার সম্পর্কে কিছু তথ্য দেয়।

গুরুত্বপূর্ণ বৈশিষ্ট্য: লেজি লার্নিং (Lazy Learning)

ইনস্ট্যান্স-বেজড লার্নিং-এর সবচেয়ে গুরুত্বপূর্ণ বৈশিষ্ট্য হলো:

- নতুন ডেটা পয়েন্ট (কোয়েরি) না আসা পর্যন্ত অ্যালগরিদমটি আসলে কিছুই করে না।
- প্রশিক্ষণ (Training) বা শেখার প্রক্রিয়াটি এখানে স্থগিত (postponed) থাকে।
- এই কারণে ইনস্ট্যান্স-বেজড লার্নিংকে লেজি লার্নিং (Lazy Learning)-ও বলা হয়। এটি কেবল ডেটা পয়েন্টগুলো ধরে রাখে।

উদাহরণ: K-Nearest Neighbors (KNN) হলো ইনস্ট্যান্স-বেজড লার্নিং-এর একটি ক্লাসিক উদাহরণ।

২. মডেল-বেজড লার্নিং (Model-Based Learning)

মডেল-বেজড লার্নিং হলো সেই পদ্ধতি, যেখানে অ্যালগরিদম ডেটার অন্তর্ভুক্ত নীতি (underlying principle) বা প্যাটার্ন খুঁজে বের করার চেষ্টা করে।

কাজের প্রক্রিয়া ও উদাহরণ

উপরে বর্ণিত একই প্লেসমেন্ট সমস্যা (আইকিউ, সিজিপিএ, প্লেসমেন্ট) মডেল-বেজড অ্যালগরিদম অন্যভাবে সমাধান করে:

1. **গাণিতিক সম্পর্ক:** যখন মডেলকে ডেটা পয়েন্ট দেওয়া হয়, তখন এটি ইনপুট এবং আউটপুটের মধ্যে একটি গাণিতিক সম্পর্ক (mathematical relationship) বা ফাংশন বের করার চেষ্টা করে।
2. **সিদ্ধান্ত সীমানা তৈরি:** এই সম্পর্কের ফলে, মডেলটি একটি সিদ্ধান্ত সীমানা (Decision Boundary) তৈরি করে (যেমন, একটি বক্ররেখা বা সমীকরণ)। এই সীমানাটি দুটি ক্লাসকে (প্লেসমেন্ট হয়েছে/হয়নি) আলাদা করে দেয়।
3. **সাধারণীকরণের নিয়ম:** এই সীমানাটিই হলো মডেলের সাধারণীকরণের নিয়ম (Generalization Rule)। মডেলটি নিয়মটি খুঁজে বের করার উপর জোর দেয়, ডেটা পয়েন্টের উপর নয়।
4. **ভবিষ্যত্বাণী:** একবার এই ডিসিশন ফাংশন (বা মডেল) তৈরি হয়ে গেলে, একটি নতুন ছাত্রের ভবিষ্যত্বাণী করার জন্য আর ট্রেনিং ডেটার প্রয়োজন হয় না। যদি নতুন পয়েন্টটি সীমানার এক দিকে পড়ে, তবে প্লেসমেন্ট হবে; অন্য দিকে পড়লে হবে না।

গুরুত্বপূর্ণ বৈশিষ্ট্য: মডেল-বেজড লার্নিং-এর লক্ষ্য হলো এমন একটি মডেল (Model) তৈরি করা, যা কিছু প্যারামিটার (যেমন, নিউরাল নেটওয়ার্কের ওজন বা লিনিয়ার রিগ্রেশনের ঢাল/ইন্টারসেপ্ট) দ্বারা সংজ্ঞায়িত হয়।

উদাহরণ: লিনিয়ার রিগ্রেশন, লজিস্টিক রিগ্রেশন, ডিসিশন ট্রি, এবং বেশিরভাগ নিউরাল নেটওয়ার্ক মডেল-বেজড লার্নিং ব্যবহার করে।

৩. ইনস্ট্যান্স-বেজড বনাম মডেল-বেজড লার্নিং: প্রধান পার্থক্য

সোর্সগুলোতে আলোচনা করা এই দুটি পদ্ধতির মূল পার্থক্যগুলো নিচে একটি তুলনামূলক আকারে দেওয়া হলো:

বৈশিষ্ট্য (Feature)

মডেল-বেজড লার্নিং (Model-Based Learning)

ইনস্ট্যান্স-বেজড লার্নিং (Instance-Based Learning)

মডেল তৈরি	প্রশিক্ষণ প্রক্রিয়ায় প্যারামিটারসহ একটি মডেল (গাণিতিক সমীকরণ) তৈরি হয়।	কোনো মডেল তৈরি হয় না; শুধু ডেটা ধরে রাখা হয়।
শেখার সময়	স্কোরিং-এর আগেই সাধারণীকরণের নিয়ম (Rule) খুঁজে বের করে প্রশিক্ষণ সম্পন্ন হয়।	শেখার বা ভবিষ্যদ্বাণীর প্রক্রিয়াটি স্থগিত (postponed) থাকে; নতুন পয়েন্ট এলে তবেই এটি কাজ করে।
ডেটার প্রয়োজনীয়তা	একবার মডেল তৈরি হলে ট্রেনিং ডেটা ফেলে দিলেও ক্ষতি নেই। ভবিষ্যদ্বাণীর জন্য শুধু মডেলটিই যথেষ্ট।	ভবিষ্যদ্বাণী করার জন্য সমস্ত ট্রেনিং ডেটা সবসময় ধরে রাখতে হয়।
সাধারণীকরণ (Generalization)	মডেলের মাধ্যমে সাধারণীকরণের একটি নিয়ম (Rule) প্রয়োগ করা হয়।	কোনো সাধারণ নিয়ম নয়, বরং ইনকামিং পয়েন্টের সাপেক্ষে ডেটাকে সরাসরি দেখে সিদ্ধান্ত নেওয়া হয়।
স্টোরেজ প্রয়োজনীয়তা	যেহেতু এটি শুধু মডেলের প্যারামিটারগুলো সংরক্ষণ করে (যেমন সমীকরণ), তাই এটি কম স্টোরেজ ব্যবহার করে।	যেহেতু এটি পুরো ডেটাসেট ধরে রাখে, তাই এটি স্টোরেজ-এর দিক থেকে ভারী (storage heavy) হয়।
অন্য নাম	-	লেজি লার্নিং (Lazy Learning)।

পরিষ্কার ধারণা:

আপনি যদি একজন ডেটা সায়েন্টিস্ট হন, তবে ভবিষ্যতে যখন কোনো অ্যালগরিদম (যেমন, লজিস্টিক রিগ্রেশন বা KNN) শিখবেন, তখন আপনার কাজ হবে চিহ্নিত করা যে সেটি মডেল-বেজড নাকি ইনস্ট্যাঙ্ক-বেজড লার্নিং করছে। এই জ্ঞান আপনাকে বুঝতে সাহায্য করবে যে অ্যালগরিদমটি কীভাবে তার সিদ্ধান্ত নিচ্ছে।

| Instance-based Vs. Model-based learning

