

CSE 471: Machine Learning

Machine Learning

Tajkia Nuri Ananna

February 22, 2023

Lecturer, Metropolitan University

Table of contents

1. Feature
2. Training, Test and Validation Data
3. Cross Validation

Feature

Feature

It is the characteristic used to uniquely identify an object.

যেমন ধরা যাক, আপনার ক্লাসে দুজন আছে আরিফ নামে। একজন চশমা পরে, আরেকজন পরে না। এখন, একদিন ক্লাসে গিয়ে আপনি আরিফকে খুঁজছেন কোনো একটি দরকারে। তাকে ক্লাসে না পেয়ে আপনি আপনার কোনো এক বন্ধুকে জিগ্যেস করলেন, ‘দোস্তু, আরিফ কোথায় রে?’ তখন সে স্বাভাবিকভাবেই আপনাকে জিগ্যেস করবে, ‘কোন আরিফ? ক্লাসে তো দুজন আরিফ আছে।’ তখন আপনি বলবেন হয় চশমা-পরা আরিফ, কিংবা চশমা-ছাড়া আরিফ।

এই যে আপনি দুজন আরিফকে তাদের একটি বৈশিষ্ট্য দিয়ে আলাদা করলেন, এই বৈশিষ্ট্যটিই হলো ফিচার (Feature)। তার মানে ফিচার হচ্ছে সেই সমস্ত অনন্য বৈশিষ্ট্য, যা কোনো কিছুকে আর দশটি সামঞ্জস্যপূর্ণ জিনিস থেকেও সহজে আলাদা করে ফেলতে পারে।

Teach the computer!

What if we want to teach a computer about Giraffes?

We will show it a picture and also include various information including the length of tail and neck, or if it has a horn or not (like the figure below)

ছবি নম্বর	লেজের দৈর্ঘ্য (সে.মি.)	গলার দৈর্ঘ্য (সে.মি.)	শিঙের মতো বস্তু কি আছে?	প্রাণীটি কি জিরাফ?
1	5	8	হ্যাঁ	হ্যাঁ
2	2	3	না	না
3	1	2	না	না
4	0	2	না	না
5	5.5	7.5	হ্যাঁ	হ্যাঁ

Feature Data

Label

Training, Test and Validation Data

Training, Test and Validation Data

Training Data

Used to train the model!

Example: Preparing for an exam.

Validation Data

Part of training data

It defines how well your model performs, how much improvement is needed or how to finetune different parameters of your model

Example: Taking the model test exams before sitting for the final exam

Test Data

Proves the performance of your trained model

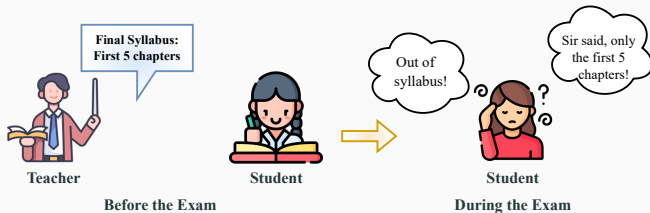
Example: The Final Exam!

Cross Validation

Suppose, Lily is a student who has her math exam coming up and there can be two distinct scenarios!

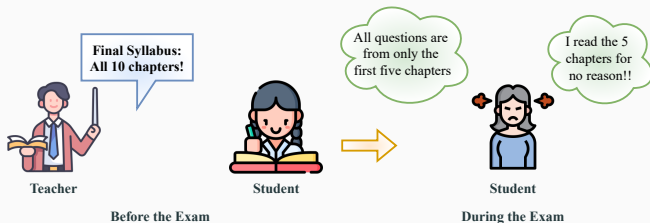
Scenario 1

Lily read only the first five chapters!



Scenario 2

Lily read all the chapters!



Cross Validation

What's the summary?

Based on the first scenario, you can't say lily is a bad student even if her exam didn't go very well!

Lily was not evaluated properly based on her preparation!

Base of Cross Validation

- The previous story is the base of Cross Validation.
- Suppose, we have a dataset 100 datapoints
- If we randomly choose the first 80 datapoints as training data and 20 datapoints as test data and our model works well we can't say our model is accurate or vv!!
- That is where cross validation comes into the rescue!

What is Cross Validation

Cross-validation is a technique in which we train our model using the subset of the data-set and then evaluate using the complementary subset of the data-set.

Process Validation Example

Process of Cross Validation

- First we divide the dataset into k numbers of equal subsets
- Each time, we train the model using $k-1$ datasets and test it using the remaining subsets.
- Finally, we calculate the average of total scores!

Example Validation Example

Example of Cross Validation

- Suppose we have a dataset with 100 points
- We divide the dataset into $k=5$ different subsets, each with 20 datapoints.

20	20	20	20	20
D1	D2	D3	D4	D5

Example Validation Example

Example of Cross Validation

- We implement cross validation in the following manner

Iteration (1 to K)	Training Set	Test Set	Performance Score
1	D2, D3, D4, D5	D1	S1
2	D1, D3, D4, D5	D2	S2
3	D1, D2, D4, D5	D3	S3
4	D1, D2, D3, D5	D4	S4
5	D1, D2, D3, D4	D5	S5

- Then finally c

