Lecture–01 1/9/24

Machine learning:

Feature Selection: অনেকগুলো col → selective cols.
always correlation বেশি col
মানেই কিন্তু optimal হবে না।

Generalization: Model predictor যদি খুলনাতে
বাসার জন্যই prediction করতে দেয়।
→ not generalized
→ overfitted for Gulshan.

Neural network → existed before 2000.
but data was scarce,
তখন এত popular হয় নাই।
computational power ও কম ছিল।
Neural networks are data hungry.
We have now data.
Data cloud
Data Generator.

- Principal Component Analysis:

  vector এর direction change করে যদি বুঝান করতে পারি যে কোনো dimension redundant হয়ে যায়, then feature/vector space কমে আসে।

  (Feature selection এ use হতে পারে)

- FNA

ML: Experience যত ↑ performance ↑

training data ↑ performance ↑

Supervised Learning:

Model build করতে Ans টাও (target Variable/label)ও provided.

— supervised learning

- **Regression**

  "আগত দাম prediction. দামও দেয়।

  ↓

  target Variable is a continuous Variable"

- **Classification**

  " Student এর level term prediction.

  Here target Variable is a discrete variable."

  - binary classification

    covid - yes/NO

  - Multi-class classification prob.

    cell - healthy/unhealthy/damaged.

Unsupervised learning

Clustering :

Target Variable বা label নাই।

Data থেকে Essence বের করতে হয়।

Reinforcement learning :

learning of human beings.

Win/loss/Rewards.

Lecture – 02 2/9/24

Label = Target Variable.

Unsupervised

Data র label নাই। We want to extract essence.

Data similarity.

Reinforcement learning:

objective function

States

Win/Loss Model.

বিভিন্ন state এ traverse করলে profit/loss analysis করে optimal এর দিকে যাওয়ার চেষ্টা করা।

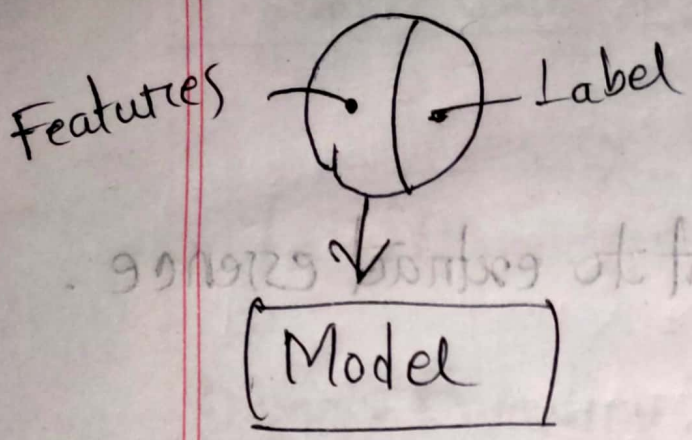
Datasets:

Supervised

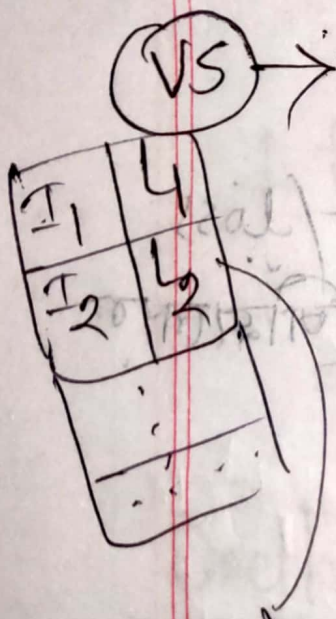
i) Training set: এতে data র সব info + label থাকবে।
Features + Labels থাকবে।

ii) Validation set: Not used in training. performance analysis এ।

TS



f(feature info) = label



Model থেকে পাওয়া

⟹ Then performance analysis করি।

Model ভালো করার উপায়:

1) Training Data ↑
2) Hypothesis, H. } House info দিলে price predict করতে পারে।
   Horizon broader করতে হবে।
3) Hyperparameter tuning.

→ Validation test ভালো না করলে Hypothesis আরও complex করতে হতে পারে।

Model, Validation Data দেখে essence নিয়ে ফেলবে। তবে performance 85% এ তেমন significant কিছু না।

Test Dataset: Model building এ কোনোভাবেই একে use করা যাবে না।

5% data নিয়ে test করলাম। Model যেখানে যেগুলো data তে ও labels দেখে নাই, তাও এই test set ফেল দিতে হবে — Ideal

paper publish → always Independent test data নেলে কেয়া হয় না।
নতুন data পাওয়া difficult
loophole থাকেই।
But industry তে নতুন data আসতেই থাকে।
So, ওখানে সোল দেয়া যেতেই পারে।

'purely unseen' — myth!

Rule of Thumb,

65 — 20 — 15

70 — 20 — 10

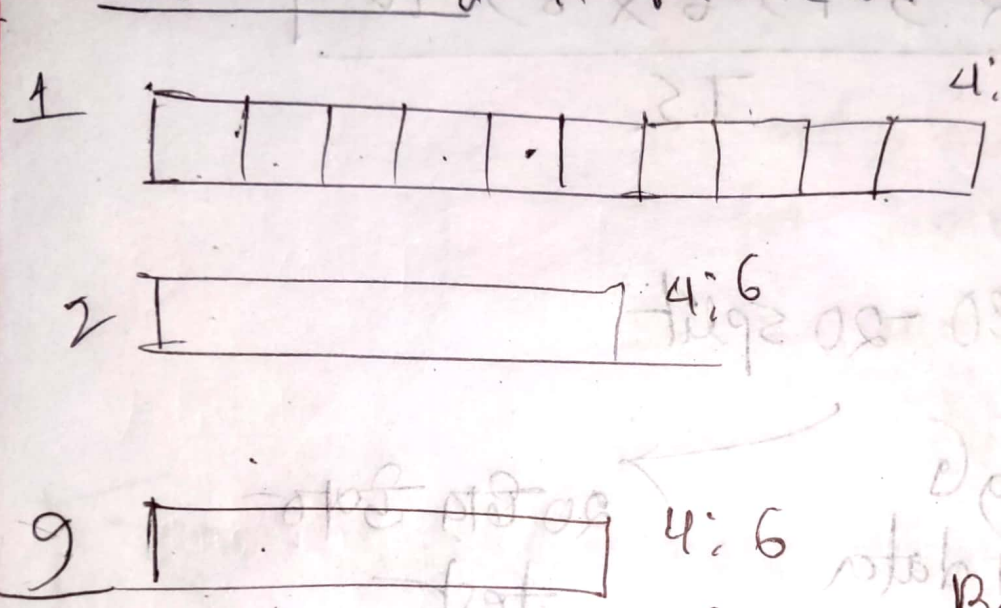
TS — VS — Test Set

একই data কীভাবে test + training এ?

Validation Test → HP. tuning

Independent Test → performance reporting.

চারকোণা আংকা: K-fold C.V.



10 টা Fold

Each has 10 data points.

90 টা data দিয়ে training করবো।

আর 10 টা Validation.

100

60% → neg

40% → pos

Balanced

Multi-class এ

যত্ন এ balance

neg - pos

50 - 50

49 - 51 ✓

0 - 1 - 2

33% - 34% - 33%

Folds | 1 2 3 4 5 6 7 8 9 | 10 |
T.S. VS

Folds 1 | 2 3 4 5 6 7 8 9 10 |
VS T.S.

pure 80 - 20 split

training ची 80 % data used होतो।

20 % data test

K-fold K = 10

90% training ची

overall 100% test data.

for hyperparameters $\langle h_1 \rangle \ldots \langle h_5 \rangle$

for K = 1 10

{

}

90% বা তার max for some h_i

So, we are so happy with h_i.

নতুন কিছু predict করতে কোন h_i use করবে?

customer কে কাকে কী দিবে?

Final Model: h_1 optimized

h_2 optimized

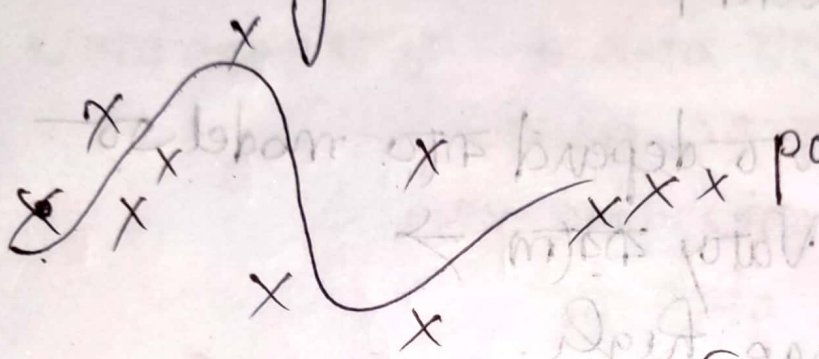
h_1, h_2 90 টা training data ← ফেলা করে (90%) ফিয়ে 2 100 টা ফিয়ে more than 90% expected

expected to be generalized.

ফিয়ে final train ফিয়ে model. train করে model ফিবে।

Lecture-3

- Hypothesis → যেই func টায় input → output mapping recover করার try করতেছি।
- Linear - Very Simple Model



যত simple, তার bias বেশি।
যে তার নিজের limit এর বাইরে বেশি কিছু চিন্তা করতে পারে না।

প্রতিটা func এর inherent bias থাকে।
ওর বাইরে ও খুব চিন্তা করতে পারে না।

- slide এ 4 টা hypothesis এরই datapoints same.
যত sophisticated / polynomial degree ↑
overfit করতে চায় তত।

- Degree 12 তে unnecessary peak দেখা যাচ্ছে।

Okham's Razor : perciminia
যেখানে সম্ভব আমি simple model এ, সেখানে Simple Model.

Bias এর opposite = Variance.

Generalized Model এ Variance থাকলো — আমার dataset এ যেমন, অনুপাত এ তেমন।

∴ sample এর উপর depend করে model এর output highly Vary করলে → Variance high.

final এ representative line পাই.

Degree-12 এর curve এর মধ্যে অনেকগুলো diff. ← Variance high.

linear এ variance low.

∴ sampling এর diff এর জন্য model output highly Vary করে → Highly Variance Model.

- Model heavily biased →
  100 টা points — 60% performance
  200 " " → 62% "
  2/00 " " → 63% "

  Variance কম → data বাড়ায় লাভ নাই।
  Model এর সীমাবদ্ধতা। Model এ যাইহ্য আর চিন্তা করতে পারবে না data যতই বাড়াও।

  We have to tradeoff bet^n bias & Variance

- Decision tree → Russell Norvig এর book এর example.

- Ordering of features to build a decision tree
  → correlation Analysis.

- Information gate

- Balanced dataset – 6টা yes, 6টা no.

→ Type যিয় যখন subsample করলাম তখনো distribution ratio same. 50-50

→ pattern যিয় check করতে 2:4 ও নামাইলাম।

1	12		
2	5	9	10

Where there is uncertainty, there is message.

$-\log p$ or $\log \frac{1}{p}$ যত probability ↑ তত info value কম ↓

Attribute property → Entropy

কোনো 1 টা Random Variable এর info content (weighted)

$$H(v) = \sum_{k} P(v_k) \log_2 \left\{ \frac{1}{P(v_k)} \right\}$$

v এর k সংখ্যক value $= \sum_{k} -P(v_k) \log_2 P(v_k)$

Entropy কতটুকু কমাইতে পারে।

Lecture -04

8/9/24

Uncertainty ↑ Statement Value ↑
probability ↓

Information is inversely proportional to probability

logarithmic

Entropy ↓ Uncertainty ↓

$$H(\text{output}) = B\left(\frac{p}{p+n}\right)$$

true আসার probability

Remainder (A)

↳ A Variable select করলে → সেই subprob গুলা পাই → যেগুলোর p আসার probability

$0 \times \log 0 = 0$
$1 \times \log 1 = 0$

} হয় না এটা guaranteed হবে " "

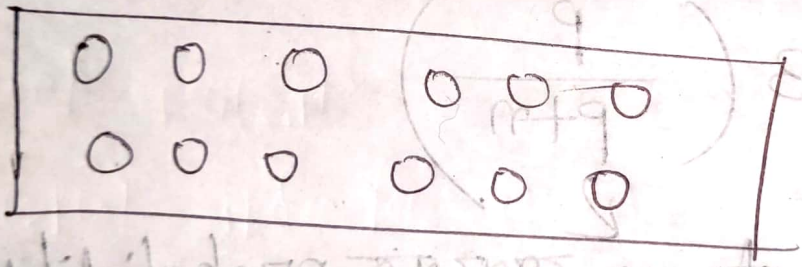
Gain(A) = B$\left(\frac{p}{p+n}\right)$ − Remainder(A)

↓ ↓

মূল prob এর entropy

prob ভাগ করলে এর entropy.

যেই A এর gain(A) ↑ বেশি ভালো।



$0.5 \log \frac{1}{2} + 0.5 \log \frac{1}{2}$

$= -1$

Feature Ranking by Information Gain.

Recursive elimination of features

Data Collection
pre-processing
Feature Engineering
Feature Selection

Domain knowledge needed to know what are important.

Performance Metrics

self-study

AUROC
AUPR
} আরও হবে

Confusion Matrix: gives stat

Row → Actual class
Col → predicted class

P = pos, N = neg

↓

অধিকাংশ dataset এমন আছে
P = ২০, dataset এ ২০ টা pos classed points আছে।

Neg class কে 0 ধরা হয়।

prediction

TP

TN

) ↳ কী predict করলাম

যা

predict করলো

আসলে সঠিক কি না

FN

↳ negative predict

false করলি

prediction,

যে সত্যিই 0 ছিল।

(T+class)

multiclass ত, diagonal বরাবর ~~অঙ্ক~~ মান।

Accuracy: পুরা dataset ত কতগুলা correct prediction করছে পারসি।

$$Acc = \frac{TP + TN}{P + N}$$

TP + FN ← total sum বরাবর

FP + TN

Model (xP)

P = 99
N = 1

Data balanced না হলে Accuracy misleading.

अन्यथा नाम Recall

S_n → positive class ची Accuracy.

$$S_n = \frac{TP}{TP + FN}$$

(sensitivity)

S_p (specifity) = neg class ची Accuracy

$$= \frac{TN}{TP + FN}$$

Data balanced ना शेल्या classwise Accuracy important.

$S_n = 70\%$ $S_p = 65\%$ ✓ good model

$S_n = 70\%$ $S_p = 40\%$ ✗ bad model.

Precision : আমি যা বলি তা precise হয়।

① আমি positive কে pos বলি।

$$Prec = \frac{TP}{TP + FP}$$

Limitation : neg নিয়ে কোনো কথা নাই।

*** Mcc

Unbalanced dataset এ ভালো accuracy ভালো measure না।

If dataset has no P point, precision 0 হয়ে যাবে।
but S_p তখন খুব ভালো হবে।
(if Model is good ofc)

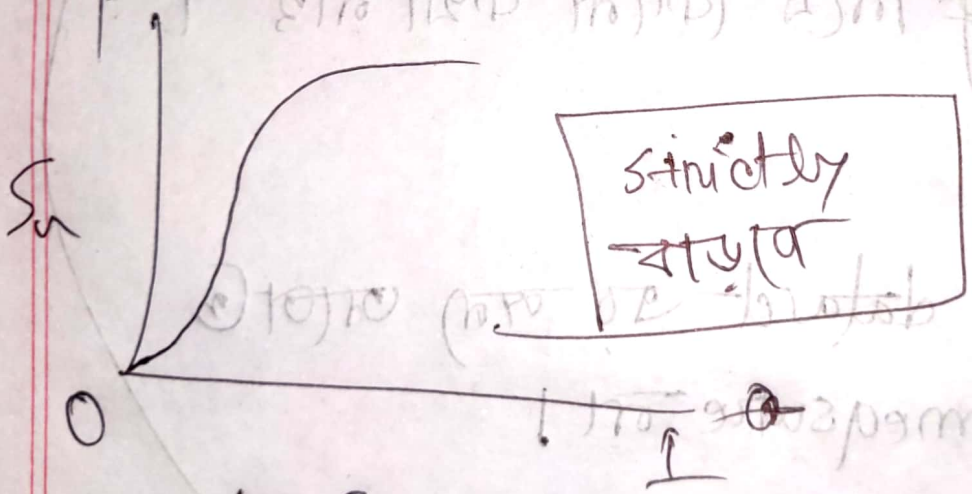
F1 Score:

Harmonic Series মনে আছে।

$$\frac{1}{\frac{\sum_{i=1}^{K} 1/x_i}{K}}$$

AUROC: / AUC

Area Under Receival Operating Curve.



যেই data balanced না, সেটায় AUROC ভালো হলেও prediction ভালো হয় না, ধোঁকা দেয়।

Linear Regression

Why linear? $w_0 + w_1 x$

$$y = c + mx$$

$$h(x) = w_0 + w_1 x$$

w_0, w_1 → known

we want to get w_0, w_1.

$$\frac{\sum_K |y - h(x)|}{K}$$

Mean absolute error বাড়াতে চাই।

$$\frac{\sum_K |y - h(x)|^2}{K}$$