Date:

Collective Wisdom.

31
3 4

Ensemble Methods

Hard vote	Soft vote

NO → Classifier 1 → Yes: 0.4, No: 0.6
Yes → " 2 → Yes: 0.9, No: 0.1
NO → " 3 → Yes: 0.4, No: 0.6

Hard vote → 1 | 2

Yes
0.4+0.9+0.4

Soft vote

0.4 + 0.9 + 0.4 = 1.7 → Win Yes

0.6 + 0.1 + 0.6 = 1.3 → No

Soft vote →

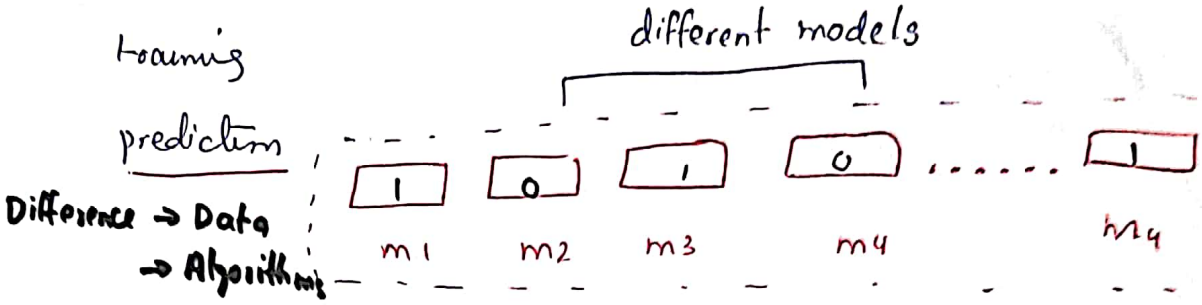
Ensemble Learning

↓

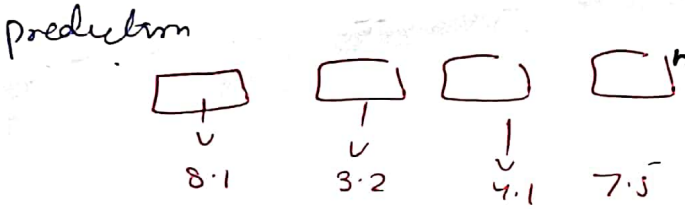
a collection of multiple machine learning models.

Wisdom of the crowd.

Core Idea



Regression :.



$$\frac{8.1+3.2+4.1+7.5}{4}$$ Mean

{ 3 → 1, 2 → 0 }

↓

majority count

Benefits

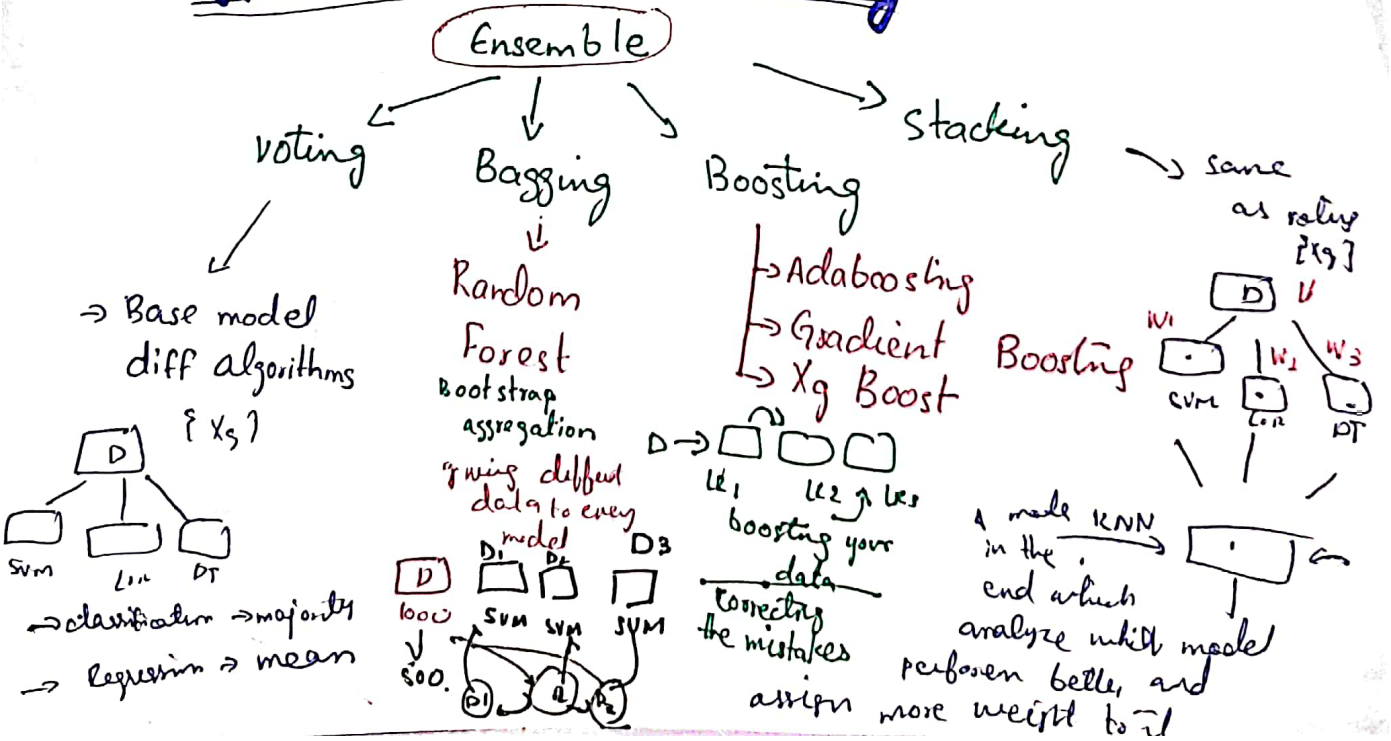
→ Improvement in performance

→ Bias variance (low bias, low variance)

→ Robustness

When to use ? always

Type of Ensemble Learning :



Random forests

↳ group of trees

↳ Bagging → Bootstrapped aggregating
↳ sampling

↓

Decision Trees

~~D_5 1000 steps~~

D_5 = 1000 rows →

DT 1 | DT 2 | DT 3 | ..... | DT 100

Sampling

↳ **row sampling** → 500 rows (random) → with replacement / without replacement

feature / **column sampling** → 5 columns (random)

→ **combination**

{new point} → prediction

DT 1 | DT 2 | DT 5 | .... | DT 100

0 1 0 1 ....

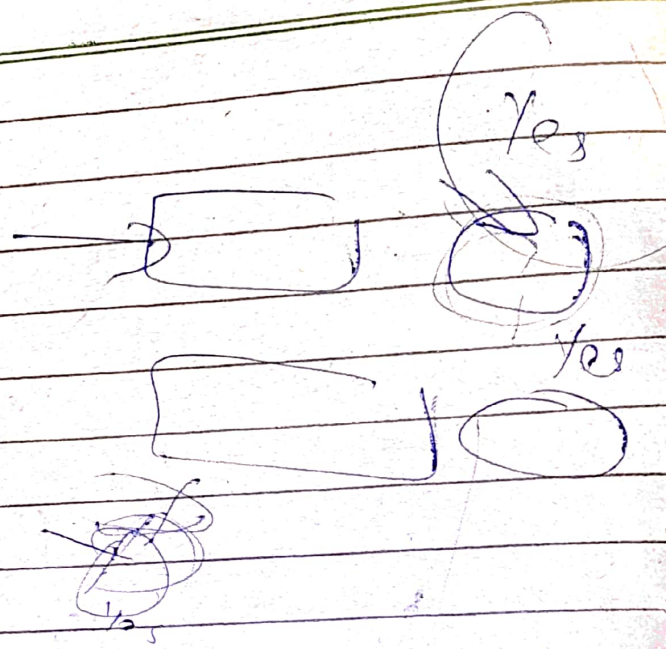
Classification → Aggregation → Most frequent

0-45 | 1-55

Regression → Mean

Day: _____ Date: _____

Bagging

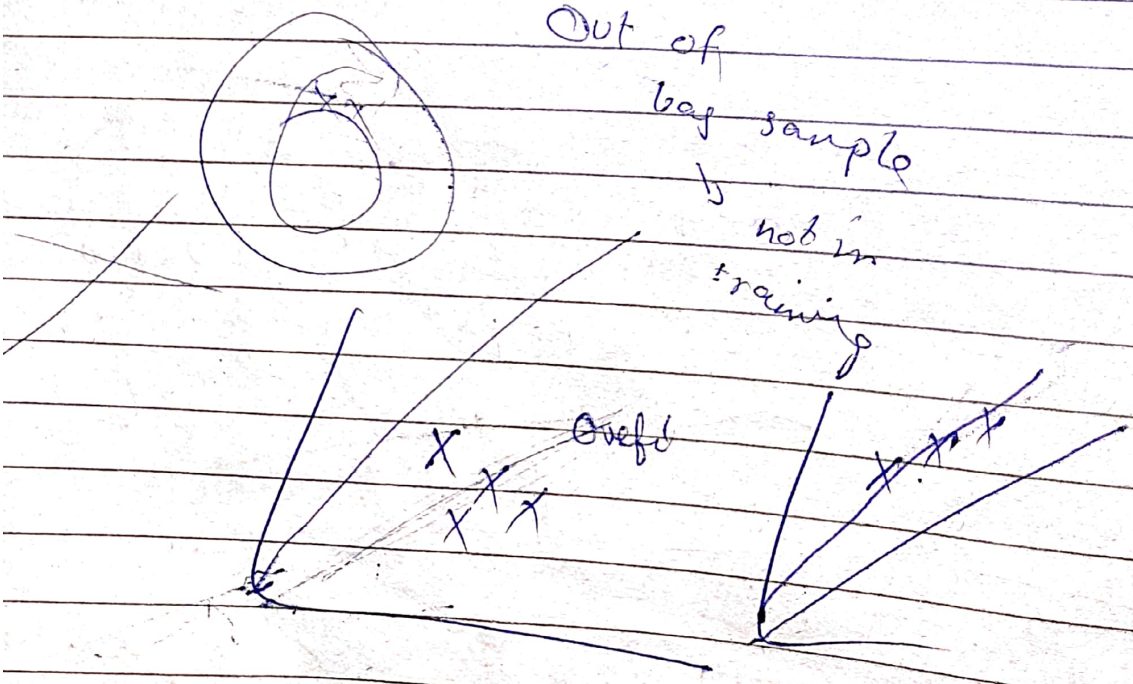


Aggregation
sab ko
collect kro

n → sample → select with replacement

train on ~~one~~ every sample

select the two



Date:

num_features = log2 m + 1

m = input features

Ensemble

Majority Voting

Bagging — Parallelism

Boosting — sequential

Stacking

Bias ↓

training accuracy ↑

Overfitting ↑

Bias ↑ T.A ↓ underfitting

Variance ↓

testing acuracy ↑

overfitting ↓

test acuracy ↓

Var ↑ overfit

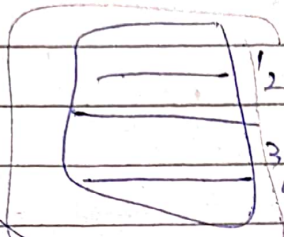
Bias ↓ T.A ↑ Test A ↓ Variance ↑

ay: Date:

Cross Validation

Validation

K = 4 K fold



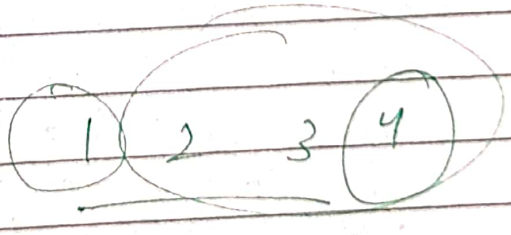
→ 1 2 3 4
training testing

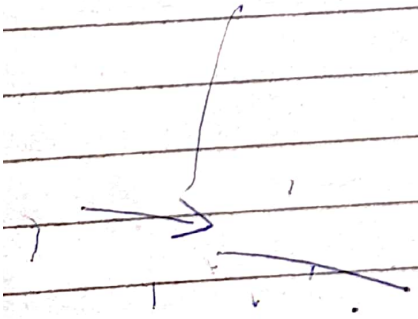
1 2 4 3
training testing

1 3 4 2
training test

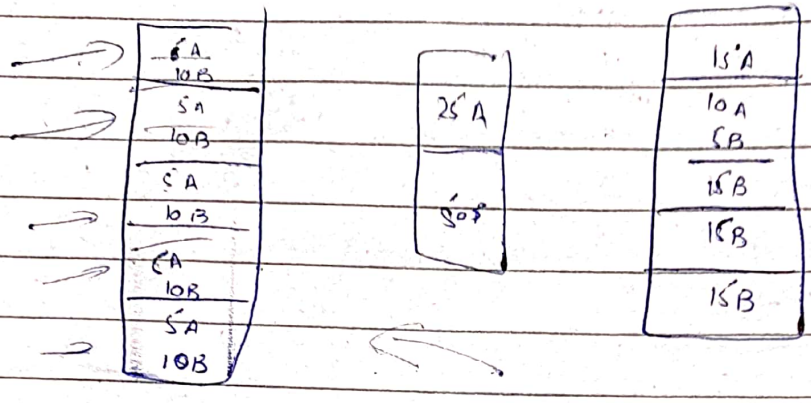
2 3 4 1
training test

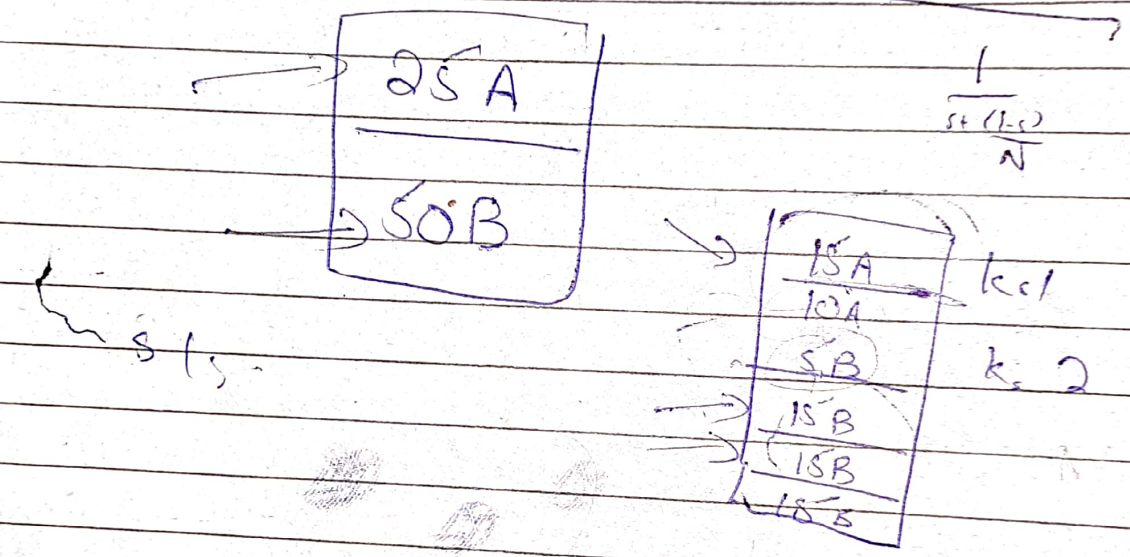
Validation: All folds average





Day:_____ Date:_____

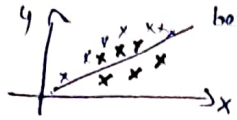




Bias and Variance

Bias → Training Data
Variance → Test data

Regression Problem



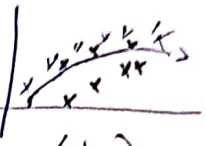
x → test data

training error high

(a)

Underfitting

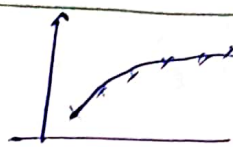
High Bias + { low variance → test data points near the line
high variance



Generalized model

(b)

less training error
low bias +
training error less than test error less } low variance



(c)

Overfitting

TE ≈ 0

Low bias + high variance

Classification Problem = {Confusion Matrix}

Model 1	Model 2	Model 3
Training Error = 2%	Train Error = 25%	Train Error < 10%
Test Error = 25%	Test error = 26%	Test error < 10%
Overfitting	Underfitting	Generalized model
Low Bias + high variance	High Bias + High Variance	Low Bias + low variance