Data Science & AI

Machine Learning

Supervised Learning

Lecture No.- 05

By- Krish Naik Sir

Recap of Previous Lecture

Topic

Topic

Topic

Topic

Topic

Regressor

Decision Tree

Probability

Topics to be Covered

Topic	cross Validation
Topic	feature transformation
Topic	Naive Baye's
Topic	Bagging
Topic	

Machine Learning

① CROSS VALIDATION ✓.

② Feature Transformation. → MinMax Scaler, → Standard Scaler }.

③ Naive Bayes } ✓.

④ Bagging And Boosting. } ⇒

① CROSS Validation And Its Types

Gate Exam

1000 datapoints

DATASET

70% → 700 TRAINING

30% → 300 TEST → Model Performance or Accuracy

Model Train → TRAINING

80% → TRAIN ⇓ Model Train

20% → Validation ⇓ Model Hyperparameter Tuning ⇒ CROSS VALIDATION

} ⇒ CROSS VALIDATION

random_state

CROSS VALIDATION

CV=5

90% ⇔ 80%

85%

	560	240		
Exp 1	TRAIN	VALIDATION	Acc 1	96%
Exp 2	TRAIN	VALIDATION	Acc 2	
Exp 3	"	"	Acc 3	
Exp 4	"	"		
Exp 5	"	"		

⇑ Average

Types of Cross Validation

① Leave One Out CROSS VALIDATION (LOOCV)

TRAINING DATA → 500 Records. {Validation data → 1}.

Disadvantage

① Time complexity is huge for training Big datasets

② Model Overfit → TRAINING ACC ↑↑ }
New Data → Validation Acc ↓↓.

② Leave p out CV

p = 10, 20, 30, 100 - - - → Hyperparameter Tuning.

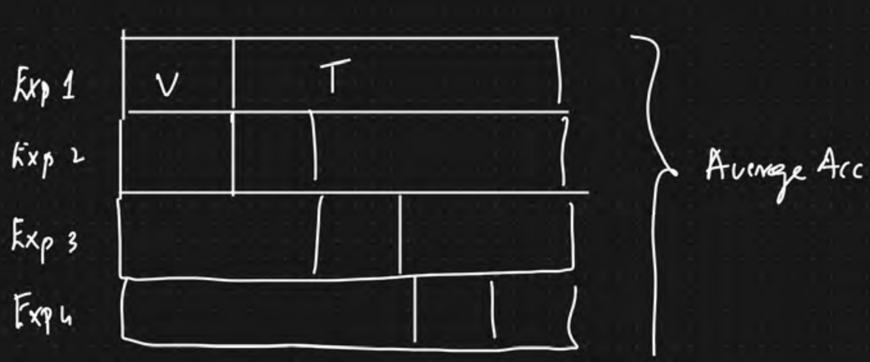
p = 50 datapoints = 500.

③ K Fold CROSS Validation

$CV = \frac{500}{100} = 5$

Validation size $= \frac{500}{5} = 100$

K = 5 n = 500



Exp

(4) Stratified K fold CV

Stratified → layers

Binary o/p

K=5 n=500 Validation = 100 ←

o/p 1's → ≈ 50%
0's → ≈ 50%
↓
Proportional representation.

n=500 100 Validation

⇒ { 350 → 1's, 150 → 0's } { 100 → 1's, 0 → 0's } { 50's ≈ 1, 50's ≈ 0 }.

(4) Time Series CV

Product Review Sentiment Analysis

Time Series Application

Product A → JAN ⟷ Dec

TRAINING VALIDATION

Day1 Day2 Day 3 Day 4 | - | Day - - Day N

Question: What is the primary purpose of using cross-validation in model building?
a) To increase the speed of the training process.
b) To reduce the complexity of the model.
c) To estimate the model's performance on an independent dataset. ✓
d) To reduce the need for feature selection.

Question: What is a special consideration when using cross-validation for time-series data?
a) Randomly dividing the dataset into k-folds.
b) Ensuring that the validation set comes after the training set in time. ✓
c) Increasing the number of folds to improve accuracy.
d) Using a larger test set than the training set.

Question: What is an advantage of cross-validation compared to using a single validation set?
a) It requires less computation time. ✗
b) It provides a more accurate estimate of out-of-sample performance. ✓
c) It eliminates the need for a test set. ✗
d) It always selects the best hyperparameters. ✓

500

i) $\boxed{b,d}$

Assignment

$K = 5, 10, 15, 20, 25,$
$50, 100, 500$

Question: In k-fold cross-validation, what is the effect of increasing 'k'?
a) Decreases both bias and variance of the model evaluation.
b) Increases bias and decreases variance of the model evaluation.
c) Decreases bias but increases variance of the model evaluation. ✓
d) Increases both bias and variance of the model evaluation.

$n = 500 \quad K = 500$

Question: In the context of cross-validation, what is the impact of partitioning data into a higher number of folds?
a) Increases the risk of overfitting. ✓
b) Decreases the size of the training set in each fold. ✓ } $\boxed{\text{A and b}}$
c) Reduces the computational complexity.
d) Makes the model less sensitive to the choice of hyperparameters.

Question: How does cross-validation help in determining the optimal complexity of a model?
a) By allowing the model to train on larger datasets.
b) Through repeated training and validation on different data subsets. ✓
c) By fixing the hyperparameters to standard values.
d) By exclusively focusing on the model's accuracy.

② Feature Scaling

(yrs) Age	(Kgs) Weight	(cms) Height
24	70	170
28	80	180
30	75	150

⇓ Transformation

[-3 to +3]

[0 to 1]

① Standard Scaling

$$SS = \frac{x_i - \bar{x}}{\sigma} \Rightarrow \underbrace{\frac{x_i - \mu}{\sigma}}_{}$$

Z-score

{Standard Normal Distribution}.

② **Min Max Scaler** ⇒ Image Application = [0 - 1].

$$\boxed{\text{Min Max} = \frac{X - X_{min}}{X_{max} - X_{min}}} \Rightarrow 0 - 1.$$

③ **Naive Baye's Algorithm**

① Probability [Independent And Dependent Events].

② Baye's Theorem

③ Naive Baye's Maths Intuition.

① **Probability**

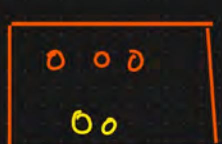
Independent Events

Rolling a dice {1, 2, 3, 4, 5, 6}.

$Pr(1) = \frac{1}{6}$ $Pr(2) = 1/6$ $Pr(3) = 1/6$.

Dependent Events

① What is the probability of first removing a orange marble and then a yellow marble?



↳① $P(O) = 3/5$ → 1st Event

↳② $P(Y/O) = 2/4$ → 2nd Events

⇓

Conditonal Probability.

$$Pr(O \text{ and } y) = P(O) * P(y/O)$$

$$= 3/5 * 2/4 = \boxed{\frac{3}{10}}.$$

$$\boxed{Pr(A \text{ and } B) = Pr(A) * P(B/A)}$$

Bayes Theorm

$$Pr(A \text{ and } B) = Pr(B \text{ and } A)$$

$$Pr(A) * Pr(B/A) = Pr(B) * Pr(A/B)$$

$$\boxed{Pr(A/B) = \frac{Pr(A) * Pr(B/A)}{Pr(B)}}$$ => Bayes Theorm.

$Pr(A/B)$ = Probability of Event A given B has occured.

$P(B/A)$ =

DATASET

x_1	x_2	x_3	O/P Yes/No
–	–	–	Yes
–	–	–	No
–	–	–	Yes
–	–	–	No

$$Pr(\overset{A}{y}/\overset{B}{(x_1, x_2, x_3)}) = \frac{Pr(y) * Pr(x_1, x_2, x_3/y)}{Pr(x_1, x_2, x_3)}.$$

$$Pr(y/(x_1,x_2,x_3)) = \frac{Pr(y) * Pr(x_1, x_2\, x_3/y)}{Pr(x_1, x_2, x_3)}$$

$$= \frac{Pr(y) * Pr(x_1/y) * Pr(x_2/y) * Pr(x_3/y)}{P(x_1) * Pr(x_2) * Pr(x_3)}$$

DATASET

x_1	x_2	x_3	O/P ⇒ (↓ Pred.)
—	—	—	Yes
—	—	—	No
—	—	—	Yes
—	—	—	No
—	—	—	Yes
—	—	—	

$y =$

$$Pr(Yes/(x_1, x_2\, x_3)) = \frac{Pr(Yes) * Pr(x_1/Yes) * Pr(x_2/Yes) * Pr(x_3/Yes)}{\cancel{Pr(x_1) * Pr(x_2) * Pr(x_3)}\ \text{Constant}}$$

$= 0.6 \Rightarrow 1.$

$$Pr(y=No/(x_1, x_2\, x_3)) = \frac{Pr(No) * Pr(x_1/No) * Pr(x_2/No) * Pr(x_3/No)}{\cancel{Pr(x_1) * Pr(x_2) * Pr(x_3)}\ \text{Constant}}$$

$= 0.4$

Naive Baye's ⇒ Probability

⊛ Ensemble Techniques [Bagging And Boosting].

Ensemble ⟶ Combining Multiple Models ⇒

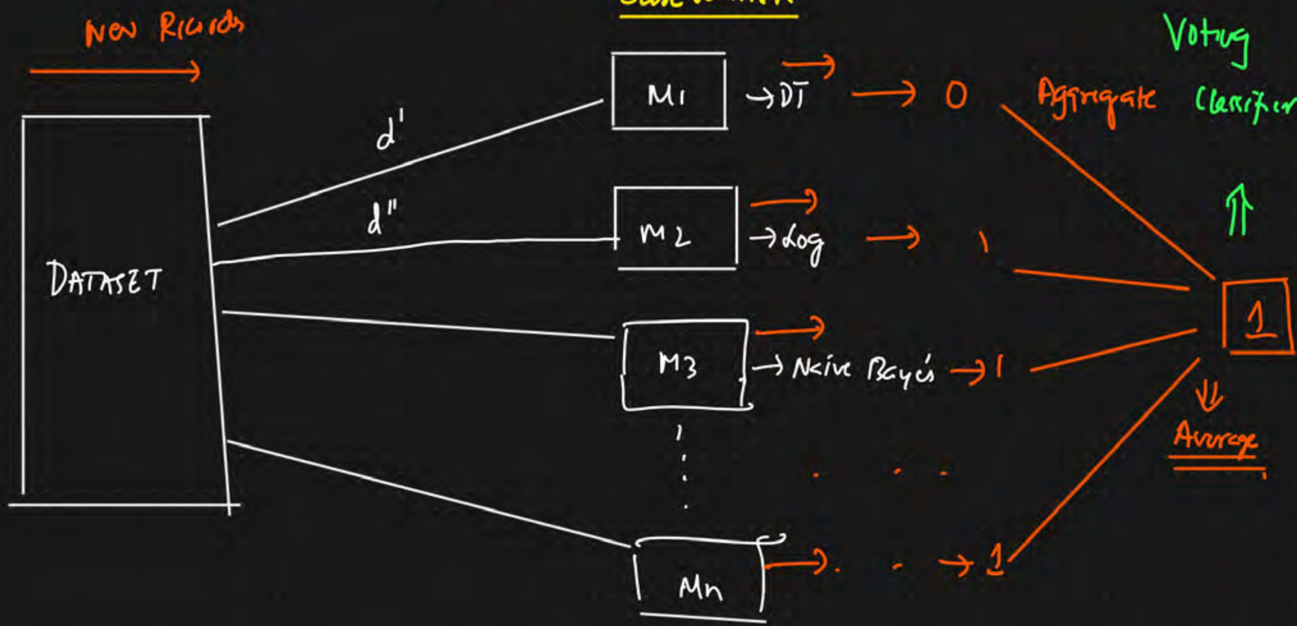
$\boxed{\text{Ensemble}}$

- Bagging
  1. Random Forest Classifier And Regressor
- Boosting
  1. AdaBoost
  2. Xgboost

Decision Tree → Overfit ⇒ Low Bias, High Variance } → Low Variance.

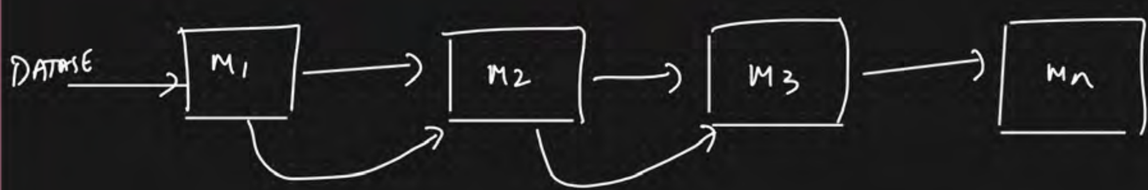
Bagging ⇒ [Bootstrap Aggregator].

Base Learners



Boosting

Weak Learner {sequentially}



Question: What is the role of Random Forests in bagging?
a) They are a type of boosting algorithm.
b) They combine multiple decision trees to reduce variance. ✓
c) They are used to sequentially correct errors.
d) They use a single decision tree with high depth.

Question: What is the primary objective of bagging in ensemble learning?
a) To combine multiple weak learners to make a final strong prediction. ✓
b) To reduce overfitting by averaging the predictions of multiple models.
c) To sequentially correct the errors of previous models.
d) To use a single, highly complex model to make predictions.

Question: How does the Naive Bayes classifier typically perform with a small amount of training data?
a) It tends to overfit easily.
b) It requires large datasets to perform well.
c) It often performs well even with a small amount of data. ✓.
d) It cannot be used with small datasets.

Naive Bayes

(1) Bernoulli Naive Bayes → Bernoulli Distribution ⇒

(2) Multinomial Naive Bayes → I/p ⇒ ~~Text~~ ⇒ Sentiment Analysis

(3) Gaussian Naive Bayes. → I/p ⇒ numerical ⇒ Guassian

Question: How is continuous data typically handled in a Naive Bayes classifier?
a) By converting it into categorical data.
b) By using the Gaussian Naive Bayes approach. ✓.
c) By ignoring continuous features.
d) By applying a linear transformation.

Question: In which of the following scenarios is the Naive Bayes classifier most effective?
a) When features are highly correlated.
b) In regression problems.
c) For large datasets with many features.
d) For text classification and spam filtering. ✓.

THANK - YOU