Machine learning

Unit - III Supervised Learning: Regression

Q.1 (a) 6M

Explain in brief techniques to reduce under fitting & over fitting.

→ 1) - The data model is cannot capture the relationship between the i/p & o/p variables accurately called underfitting.
- It generate a high error rate on both the training set & unseen data.
- Underfitting destroy the accuracy of our ML model.
- It occurs when a model is to simple, which can be a result of a model needing more training time, more i/p features.
- Underfitting can be avoided by using more data & also reducing the features by feature selection.
- e.g - building a linear regression model over non-linear data.

• Techniques to reduce underfitting.
1. Increase model complexity
2. Increase number of features, performing feature engineering
3. Remove noise from the data
4. Increase the duration of training to get better result
5. High bias & low variance

2) Overfitting
- Overfitting is a opposite of underfitting.
- It occuring when the model has been overtrained or when it contains too much complexity, resulting high error rates on test data
- Overfitting is the main problem that occur in supervised learning
- When a model perform very well for training data but has poor performance for test data is known as overfitting.
- Reasons for overfitting - 1. High variance & low bias
2. The model is too complex
3. The size of training data
- Decision tree is always in overfits.

- Techniques to reduce overfitting
  1. Increase training data
  2. Reduce model complexity
  3. Early stopping during the training phase
  4. Ridge Regularization & lasso Regularization
  5. Use dropout for neural networks to tackle overfitting

(b) Find the equation of linear regression line using following data: 6M

X	Y
1	3
2	4
3	5
4	7

X	Y	XY	X^2
1	3	3	1
2	4	8	4
3	5	15	9
4	7	28	16
$\Sigma x = 10$	$\Sigma Y = 19$	$\Sigma XY = 54$	$\Sigma x^2 = 30$

- The equation of regression line is

$$Y' = ax + b$$

$$a = \frac{n\Sigma XY - \Sigma x \Sigma Y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{4 \times 54 - 10 \times 19}{4 \times 30 - 100}$$

$$= \frac{216 - 190}{120 - 100}$$

$$= \frac{26}{20}$$

$$\boxed{a = 1.3}$$

$$b = \frac{1}{n}(\Sigma Y - a \times \Sigma X)$$

$$= \frac{1}{4}(19 - 1.3 \times 10)$$

$$= \frac{6}{4}$$

$$b = 1.5$$

$$Y' = aX + b$$

$$Y' = 1.3X + 1.5$$

For $X = 1$ $\quad Y = 1.3 \times 1 + 1.5 = 2.8$

For $X = 2$ $\quad Y = 1.3 \times 2 + 1.5 = 4.1$

For $X = 3$ $\quad Y = 1.3 \times 3 + 1.5 = 5.4$

For $X = 4$ $\quad Y = 1.3 \times 4 + 1.5 = 6.7$

(c) Write short note on:

① MAE

② RMSE

③ R^2.

→ 1) MAE

- MAE stand for Mean Absolute Error.
- MAE is a very simple metric.
- MAE calculates the absolute difference bet" actual & predicted values.
- Example:- We have i/p data & o/p data & use Linear Regression which draw a best fit line.

Actual o/p ↓ ↓ predicted o/p.

- $MAE = \frac{1}{N} \Sigma |Y - \hat{Y}|$

So, sum of all the errors & divide them by a total number of observation to get a MAE.

• Advantage
① The MAE get same unit as the o/p variable
② It is most robust to outliers

• Disadvantage:
① The graph of MAE is not differentiable

2) RMSE
- RMSE stands for Root Mean Squared Error.
- It measures the average difference betⁿ a statistical model's predicted values & the actual values.
- RMSE is a simple square root of mean squared error (MSE).

$$\therefore \quad RMSE = \sqrt{MSE}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)}$$

• Advantages:-
① It gives a relatively high weight to large errors

• Disadvantage:-
① Not robust to outliers

3) R^2
- R^2 is also known as coefficient of determination or sometimes also known as godness of fit
- R^2 score is used to evaluate the performance of a linear regression model.

$$R^2 \text{ squared} = 1 - \frac{SS_r}{SS_m}$$

SS_r = Squared sum error of regression line
SS_m = squared sum error of mean line

6M Regularization under the Lasso & ridge regression

Q.2.(a) Explain in brief lasso & Ridge Regression

→ 1) Lasso Regression
① In machine learning, Lasso stands for least Absolute Shrinkage & selection operator.

② It is a regression analysis method.
③ lasso is used for the overcome of overfitting.
④ Lasso regression is a regularization technique
⑤ It uses shrinkage [shrinkage → Reduction]
⑥ If a regression model uses the L1 regularization technique, it is called Lasso Regression
⑦ Lasso regularization has the ability to set some coefficients to zero.
⑧ Lasso applied in economics & finance
⑨ It is found to improve prediction.

2) Ridge Regression
① Ridge regression is a model tuning method
② It is used to analyse any data that suffer from multicollinearity
③ Multicollinearity :- It is a statistical concept where several independent variables in a model are co-related.
④ This method performs L2 regularization
⑤ Ridge regression model can be used on dataset that have many correlated features.
⑥ Ridge regression is mostly used to reduce the overfitting in the model
⑦ It reduce the complexity of the model
⑧ Ridge regression is a type part of linear regression

(b) what is bias & variance trade off for ML models
→ [ex. first explain ridge regression]
① In machine learning, if we minimize one component of the error [i.e bias], & the other component [i.e variance] tends to increase & vice versa
② finding the right balance of bias & variance is key to creating an effective & accurate model called as bias - variance tradeoff

③ The bias-variance tradeoff describes the relationship betn a model's complexity.

④ Applications :-

ⓐ Regression ⓑ classification ⓒ Reinforcement learning ⓓ Human learning.

⑤ Diagram

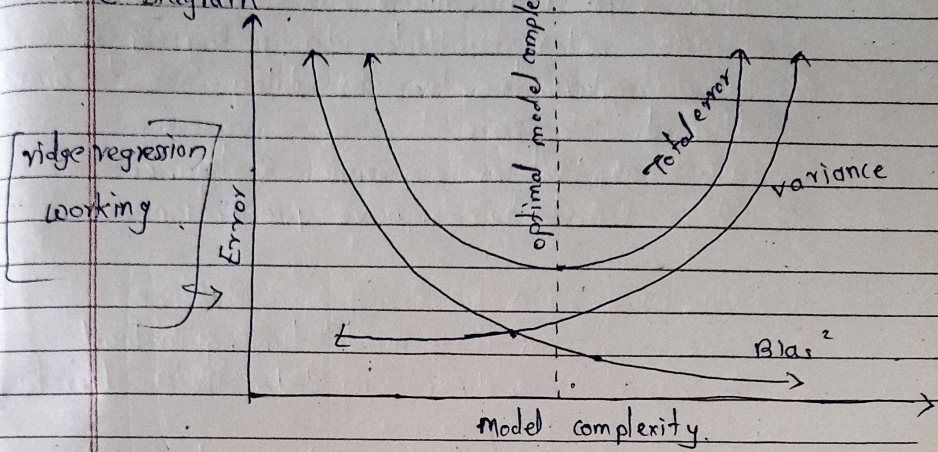


Fig - Bias-variance tradeoff

ⓒ Write short note on Evaluation metrics.

→ • Evaluation metrics.

① Evaluation metrics are used to measure the quality of the statistical or machine learning model.

② There are many different types of evaluation metrics available to test a model.

③ The evalution metrics includes classificaton accuracy, logarithmic loss, confusion matrix & so on.

④ classification accuracy is the ratio of the number of correct predictions to the total number of input samples.

⑤ Logarithmic loss also called log loss.

⑥ A confusion matrix gives us a matrix as o/p & describe the complete performance of the model

• Applications of Evaluation metric.
① statistical analysis
② Machine learning

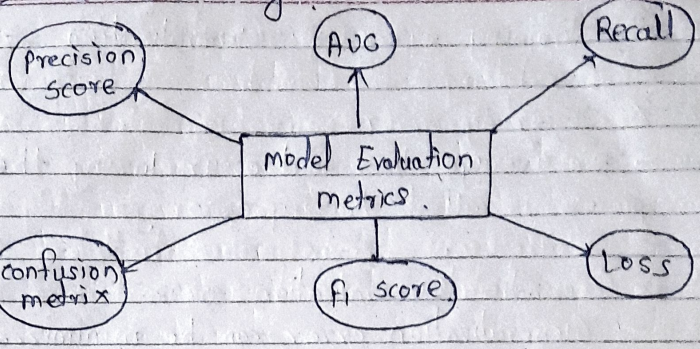


Fig:- model Evaluation metrics

[* Evaluation metrics :- MAE, RMSE, R^2]

Linear Regression

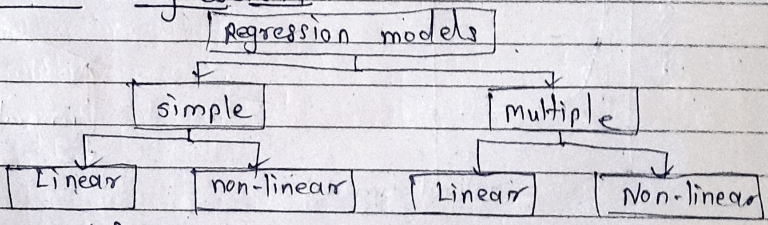


fig. Types of regression model.

① It is used to predict the value of variable based on the value of another variable
② It is a data analysis technique, that predicts the value of unknown data by using another related & known data value
③ It provide continuous output.
④ It is used to solve regression problem
⑤ It is used to draw best fit line
⑥ It is most popular machine learning algorithm.
⑦ α & β are parameters of regression
⑧ $y = mx + c$ where y = dependent variable
x = independent variable m = slope of line
c = estimate intercept.

⑨ It provide a scientific calculations for identifying & predicting future outcomes.

* ② Training error & Generalization error.

1) Training error.
- The average loss that occur during the training process is called training error.

2) Generalization error
- Generalization error can be minimized by avoiding overfitting in the learning algorithm.
- Types of generalization error
  ① Bias ② variance, ③ irreducible error.

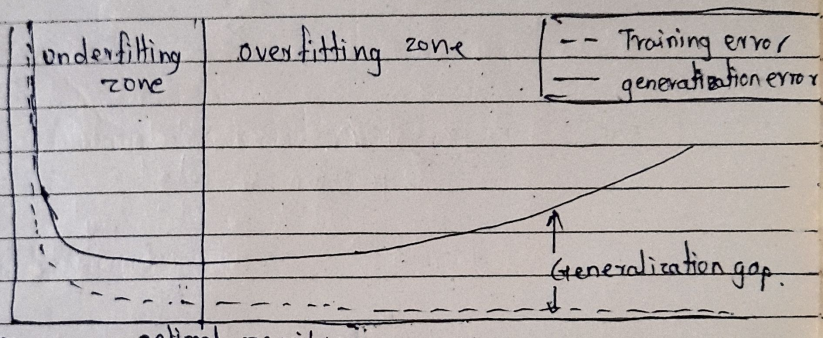


fig. Training error & generalization error.

* The Gradient Descent algorithm
① It is an optimisation algorithm which is commonly used to train machine learning models & neural n/ws.
② The GD procedures is an algorithm for finding the minimum of a function.
③ This method is commonly used in ML & DL to minimize a cost function.
④ GD algorithm does not work for all functions. There are 2 specific requirement a function has ① Differentiable ② convex.

- GD method's steps

1. choose a starting point (initialization)
2. Calculate gradient at this point.
3. make a scaled step in the opposite direction to the gradient.
4. repeat point 2 & 3 until one of the criteria is met:
   - i) maximum no. of iterations reached.
   - ii) step size is smaller than the tolerance.

Parameter tunning
trial & error.

① Bias
② variance
③ Underfitting
④ overfitting
⑤ Generalise.

Date Page

under
1) Overfitting - high Low bias & Low variance
2) Underfitting - over low bias & high variance.

① Regression - performance matrix
② classification - confusion matrix. - precision or recall.

categorical data

Low bias - error in training data
variance - error in testing dataset.
testing

Generalize - equal bias & variance.

— Always overfit in decision tree

L2 =
* Ridge regression = Loss + $\alpha \|w\|^2$ (penalty)
[same feature & same value multicollinearity problem solve ridge regression use ridge]

— It minimize the complexity of problem
- penalty is reduce the loss. [penalty should be change]
- solve overfitting problem

$\|w\|$ - vector of coefficient, α = constant

— if α increases, the magnitude of coefficient will be reduce. is almost equal to 0

L1 =
* Lasso regression = Loss + $\alpha \|w\|$
- if α incre. coefficient is almost zero but not 0
- Reduce the loss.
- solve overfitting problem.

Loss is calculate by using cost function.

Unit 4 - Supervised Learning: Classification

Q.1 Explain in brief methods used for evaluating classification models.

SM

→
1. Accuracy
2. Precision
3. Error rate
4. Recall (sensitivity)
5. specificity
6. f1 score
7. AUC - ROC
8. Log Loss.

1) Accuracy
- Accuracy is the ratio of the number of correct predictions and the total number of predictions
- Accuracy is useful for balanced classes but not good choice for unbalanced classes
- formula :

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

2) Precision
- Precision is a measure of exactness
- The ability of a classification model to identify only the relevent data points.
- formula :

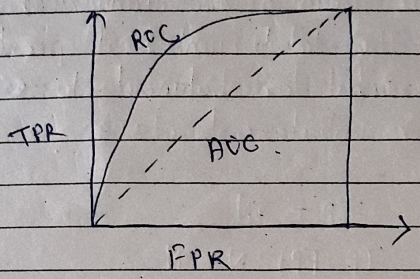
$$\text{precision} = \frac{TP}{TP + FP}$$

3) Recall -
- Recall is a measure of completeness.
- ~~The~~ It has ability to detect positive samples.
- Recall & sensitivity are similar.
- formula

$$\text{Recall} = \frac{TP}{TP + FN}$$

4) AUC - ROC

- AUC stands for Area under the Curve & ROC stands for Receiver Operating characteristics.
- ROC is the probability curve & AUC represents the degree or measure of separability.
- It tells how much the model is capable for distinguishing betn classes.



① TPR (True Positive Rate) (Recall).

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

② FPR (False positive Rate)

$$FPR = 1 - Recall$$

$$= \frac{FP}{TN+FP}$$

5) Log loss.

- It is also called as Logarithmic loss or logistic regression.
- Logg loss measures the quality of probabilistics predictions, f1 score ignores the probabilistic ~~pr measure of~~ nature of classification.
- It takes input as a probability value between 0 & 1.
- Log loss is closer to '0'.
- A lower log loss value means better predictions.

Q.2 Consider the following data to predict the student pass or fail using the K-Nearest Algorithm (KNN) for the values physis = 6 marks, chemistry = 8 marks with number of neighbours k = 3

Physics (marks)	chemistry (marks)	Results
4	3	Fail
6	7	Pass
7	8	Pass
5	5	Fail
8	8	Pass

→ Physics = 6 marks, chemistry = 8 marks

k = 3

- According to Euclidean distance

$$d = \sqrt{\sum_{i=1}^{k} (x_{observed} - x_{actual})^2}$$

Here,

we have 2 samples,

so,

$$d = \sqrt{(x_{O1} - x_{A1})^2 + (x_{O2} - x_{A2})^2}$$

① $\sqrt{(6-4)^2 + (8-3)^2} = \sqrt{29} = 5.38$

② $\sqrt{(6-6)^2 + (8-7)^2} = (1)$

③ $\sqrt{(6-7)^2 + (8-8)^2} = (1)$

④ $\sqrt{(6-5)^2 + (8-5)^2} = \sqrt{10} = 3.16$

⑤ $\sqrt{(6-8)^2 + (8-8)^2} = (2)$

out of this select '3' nearest value because the value of k = 3

out of these select '3' nearest value because the value of K=3
i.e 1, 1, 2
- Nearest selected values have "pass result" and zero fail
i.e 3 > 0
so the student is Pass

Q.3 Write short note on Ensemble learning methods
i) Simple (mean, mode, average) ii) Advanced (bagging, boosting) OR
Q: What do you mean ~~by~~ ensemble learning & difference betn bagging & boosting
- Types of Ensemble learning.

→ Ensemble learning – It is a technique that combines multiple ML algorithms to produce one optimal solution
- It ~~improves~~ helps to improve ML results & decision
- In learning models noise, variance, bias are major sources of error these methods helps to minimize these errors

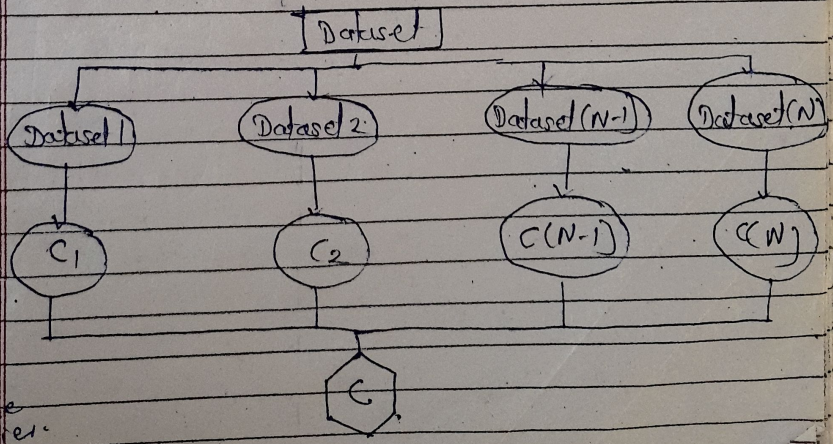


fig. An overview of Ensemble learning.

1) Simple :-
① mean / average
② mode
③ The weighted average

1) mean / average.
- In the mean/average ensemble techniques, data analyst takes average prediction made by all models when making ultimate predictions.

2) Mode.
- In statistical terminology, 'mode' is the no. or value that most appears in a dataset
- In ensemble learning techniques, ML professionals use a no. of models for making predictions about each data point.

3) The weighted average.
- In the weighted average ensemble method, data scientists assign different weights to all the models in order to make a prediction.

• Types of ensemble

2) Advanced :- ① Bagging ② boosting ③ Adaboost ④ RF

	Bagging	Boosting
purpose	① Reduce variance	① Reduce bias
memory use	② High	② Low
weight	③ Each model receives equal weight	③ models are weighted according to their performance
base model training	④ Parallel	④ sequential
CLP used	⑤ High	⑤ low
sampling method	⑥ Random sampling	⑥ Systematic sampling
computing intensive	⑦ Highly computing intensive	⑦ Less computing intensive
Example	⑧ Random forest	⑧ Adaboost
Robust	⑨ Robust	⑨ Not robust

6M

③ Adaboost. & ④ Random forest.

* Adaboost.

- Adaboost stands for Adaptive boosting
- It is a popular boosting algorithm
- Adaboost algorithms can be used for both classification & regression problem.
- Adaboost helps you combine multiple "weak classifier" into a single "strong classifier".
- It is indirectly proportional to the weak learners.

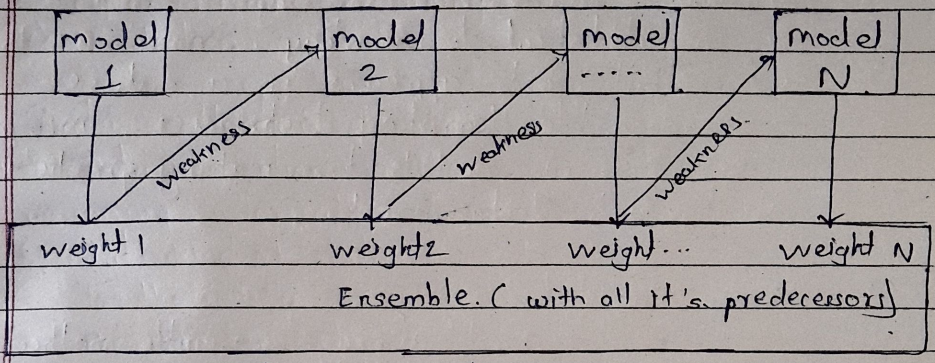


fig. AdaBoost Algorithm.

④ Random forest.

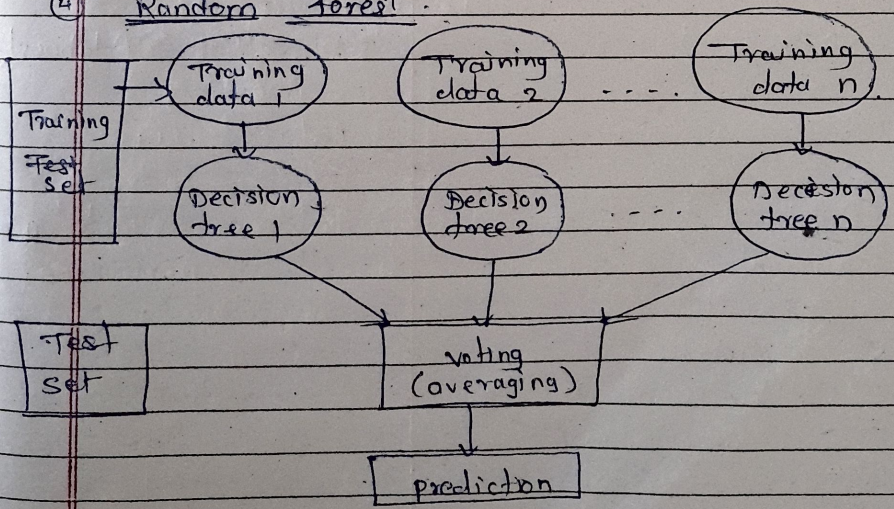


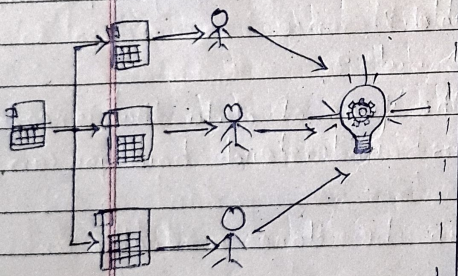
fig. Working of Random forest algorithm.

Date
Page

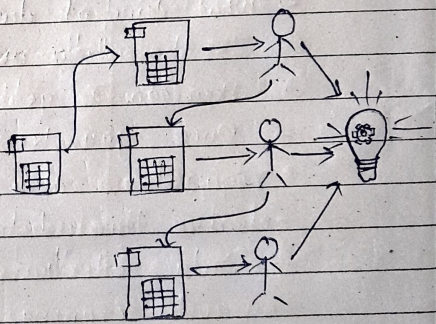
+ Random forest is a popular ML algorithm that belongs to the supervised learning technique.
- It can be used for both classification & regression problem in ML.
- It good predict produce good prediction than can be understand easily.
- Random forest is a classifier that contains a number of decision trees of the given dataset.
- It is used as black box model in business.

• Working of ~~baggi~~ Random forest algo.
- Ensemble use ~~of~~ two types of methods
① Bagging } Explain difference.
② Boosting

Bagging



Boosting



• Advantages.
① It takes less training time as compare to other.
② It used in classification & regression problem.
③ It solve overfitting problem.
④ It predicts the o/p with high accuracy.
⑤ It can handle large dataset efficiently.
⑥ It automatically handle missing values.

• Disadvantages
① Random forest is highly complex.
② It can take lot of ~~of~~ memory.
③ More trees slow down the model.
④ Random forest is much noisy.

Q. Write short note on importance of confusion matrix. 6M

OR Steps in ML modeling

⇒ • Steps in ML modeling

1) Data collection
2) Data preprocessing
3) Model selection
4) Model training - Training & testing, k-fold cross valid°
5) Model Evaluation - ① Accuracy ② Precision ③ Recall ④ Confusion matrix
6) Hyper Parameter Tuning
7) Prediction

1) Data collection

- Data collection is the process of gathering & measuring information from countless different sources.

① What kind of data is needed to solve given problem?
② Is the data available?
③ How can I get the data?

- Once you know the type of data that is ~~need~~ required, you must understand how you can derive this data.
- Data collection can be done manually or by web scraping.

2) Data preprocessing

- It is a technique that is used to convert the raw data into clean dataset.

• Data preprocessing steps -

① Removing duplicates - It is an important part of data analysis it analyzing duplicate values & remove them.
• Pandas drop-~~unders~~duplicate() method helps in removing duplicate value from dataframe.

② Transformation of data using function or mapping
- Data transformation changes the format, structure or value of the data & converts them into clean, usable data.

③ Replacing values
- Pandas dataframe replace() function is used to replace a string, regex, list, dictionary, series, numbers, etc from a dataframe.

④ Handling missing ~~ve~~ data
- Missing values can be handled by deleting the rows or columns having null values
- isna() & isnull() function used to check missing data.

3) Model selection.
- model selection is a procedure that select the most suitable model for dataset.
- Common techniques for model selection includes cross validation, hyperparameter tunning & using evaluation metrics.
- Choose appropriate ML algorithm based on the nature of your problem. (e.g. classification, regression, clustering).
- Select a set of candidate models if necessary & consider their strength & weaknesses.

4) Model training
- In ML data preprocessing, we divide our dataset into training set & Test set
- This is an important step because it improve the performance of our ML model.
① Training set - we already know the o/p.
② Testing set - by using the test ~~te~~ set model predict the o/p.

• K-fold cross validation
- cross-validation is a resampling procedure
- It is a technique used to determine/check the performance of ML model.
- It involves splitting the dataset into "k" subsets or "folds".

5) Model Evaluation
① Precision, ② recall ③ Accuracy
④ Confusion matrix
- A confusion matrix is a table that shows how well a ML model is performing.
- It compares the models prediction with the actual labels in the datasets.
- The matrix has four ~~matrix~~ values:
① True positives ② True negatives
③ False positives ④ False negatives.
- It calculate the difference betn actual value & predicted value.

• Importance of confusion matrix.
- The confusion matrix is important in ML it helps us evaluate how well our model is performing by showing us the no. of true positives, true negatives, false positives & false negatives.
- With this information, we can calculate metrics like accuracy, precision, recall & F1 score.
- Its like a road map to understand the strength & weaknesses of our models predictions.

6) Hyper parameter tuning - [It is an important step in model development proces.]
- Hyperparameter tuning in ML model is all about finding the best settings for the model.
- By adjusting things like learning rate, regularization strength or no: of hidden units that is performing our model best

7) Predictions :-
- Once the model is evaluated & improved, it is finally used to make prediction.
- The final o/p can be a categorical variable or it can be a continuous quantity.

Q Explain with example the variant of SVM, the support vector regression. 5M

OR

Define following terms with reference to SVM
(i) Separating hyperplane (ii) Margin. 6M

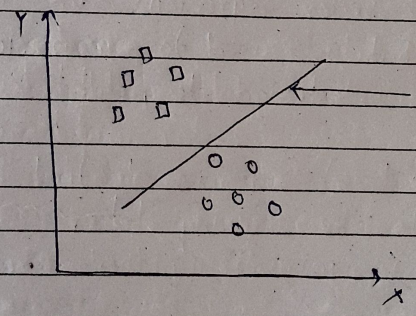
⇒ SVM - Support vector Machine
- SVM is a type of supervised learning that can be used for classification & regression
- Support vector regression (SVR) is a variant of SVM, that is used for regression task
- SVR is design to predict continuous numeric value.
- The goal of SVR is to find a function that best fits the training data.
- SVR is minimizing the error.

Example :- Consider we have a dataset of houses with the following features & prices.

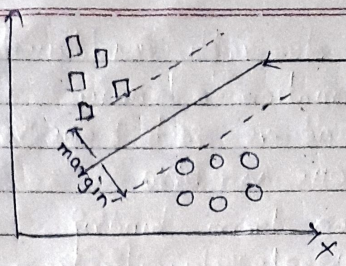
Square footage	No. of bedrooms	Location	Price
1500	3	1	200,000
2000	4	2	3,00,000
1800	3	1	2,50,000
2200	4	3	3,50,000
2500	5	2	4,00,000

Using SVR, we can train a model on this data to predict house prices based on the feature.

1) Separating Hyperplane
- Separating hyperplane is the line which is used to separate the dataset
- If data is n dimensional we required n-1 dimensional hyperplane
- SVM algorithm helps to find the best line or decision boundary, this best boundary is called as hyperplane
- It helps to determine which datapoint belong belongs to which class
- It helps to improve the generalization & robustness of the model
-

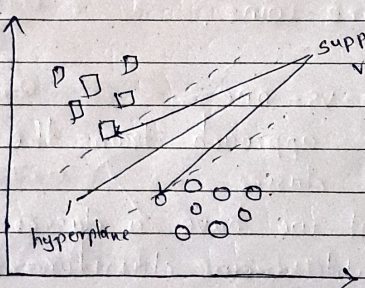


2) Margin
- The distance between the vectors & the hyperplane is called as margin.
- The hyperplane with maximum margin called optimal hyperplane.
- we always choose the larger margin because if any mistake made in model then it can be easily solve it.



- It allows better classification & generalization.

3) Support vector
- Support vector is a data point that is closest to the decision boundary (hyperplane).
- SVM uses ~~optimal hyperplane~~ to create an optimal hyperplane that maximizes margin to improve performance of model.
support vectors



Q what are different variants of multi-class classification? Explain them with suitable example.

- In ML multi-class classification is the problem of classifying data points into one of three or more classes.
- Many classification algorithm can be used to turn binary algorithm into multinomial classifier such as multinomial logistic regression

- Variants of multi-class classification
1) transformation to binary
   (a) one-vs-rest [one-vs-all]
   (b) one vs-one
2) extension from binary
3) hierarchical classification

1) transformation to binary
(a) one-vs-rest [one vs-all]
- In multi-class problem, we create a separate binary classifier for each class.
- For example, if you have classes A, B & C, you create three binary classifier: A vs {B, C}, B vs. {A, C}, and C vs. {A, B}.
- The classifier with the highest probability or score wins.

(b) one-vs-one
- This approach trains the classifier on each pair of classes.
- eg if we have a classes A, B & C it would train classifiers for A vs B, A vs C & B vs C.
- The classifier with the highest score wins.

2) ~~Extension from~~ binary
- Example of multi-class classification
1. Image classification
2. speech recognization
3. Sentiment analysis
4. Handwriting recognization
5. Disease diagnosis
6.

Q. Calculate macro average precision, macro average recall & macro average F-score for the following given confusion matrix of multi-class classification

Actual values ↓ \ Predictions →	A	B	C	D
A	100	80	10	16
B	0	9	0	1
C	0	1	8	1
D	0	1	0	9

Macro

⟹ 1) Macro average precision

- First find out the precision for A, B, C & D.

(1) ~~P(A)~~

$$\text{Precision} = \frac{TP}{TP+FP}$$

(1) $P(A) = \frac{100}{100+0+0+0} = \frac{100}{100} = 1$

(2) $P(B) = \frac{9}{9+80+1+1} = \frac{9}{91} = 0.099$

(3) $P(C) = \frac{8}{8+10+0+0} = \frac{8}{18} = 0.444$

(4) $P(D) = \frac{9}{9+16+1+1} = \frac{9}{27} = 0.429$

$$\text{Macro average precision} = \frac{P_1+P_2+\dots+P_n}{n}$$

$$= \frac{1+0.099+0.444+0.429}{4}$$

$$= \frac{1.972}{4} = 0.493$$

2) Macro average recall

$$\text{recall} = \frac{TP}{TP+FN}$$

$$R(A) = \frac{100}{100+80+10+10} = \frac{100}{200} = \frac{1}{2} = 0.5$$

$$R(B) = \frac{9}{9+0+0+1} = \frac{9}{10} = 0.9$$

$$R(C) = \frac{8}{8+1+1+0} = \frac{8}{10} = 0.8$$

$$R(D) = \frac{9}{9+0+1+0} = \frac{9}{10} = 0.9$$

$$\text{Macro average recall} = \frac{R_1 + R_2 + \dots + R_n}{n}$$

$$= \frac{0.5+0.9+0.8+0.9}{4}$$

$$= \frac{3.1}{4}$$

$$= 0.775$$

3) Macro average F1-score

$$F_1\text{-score} = \frac{2 \times \text{precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP+FP+FN}$$

$$F1(A) = \frac{2 \times 1 \times 0.5}{1+0.5} = \frac{1}{1.5} = 0.666$$

$$F_1(B) = \frac{2 \times 0.099 \times 0.9}{0.099+0.9} = \frac{0.1782}{0.999} = 0.178$$

$$F_1(C) = \frac{2 \times 0.444 \times 0.8}{0.444 \times 0.8} = \frac{0.7104}{1.244} = 0.571$$

$$F_1(D) = \frac{2 \times 0.429 \times 0.9}{0.429 \times 0.9} = \frac{0.772}{1.329} = 0.581$$

Macro average f_1-score = $\frac{f_{1(1)} + f_{1(2)} + f_{1(3)} + f_{1(4)}}{n}$

$= \frac{0.666 + 0.178 + 0.571 + 0.581}{4}$

$= \frac{1.9963}{4}$

$= 0.499075$

Micro

1) Micro average precision :-

$= \frac{TP(A) + TP(B) + TP(C) + TP(D)}{TP(A) + TP(B) + TP(C) + TP(D) + FP(A) + FP(B) + FP(C) + FP(D)}$

$= \frac{100 + 9 + 8 + 9}{100 + 9 + 8 + 9 + 0 + 82 + 10 + 12}$

$= \frac{126}{230}$

$= 0.548$

2) Micro average recall :-

$= \frac{TP(A) + TP(B) + TP(C) + TP(D)}{TP(A) + TP(B) + TP(C) + TP(D) + FN(A) + FN(B) + FN(C) + FN(D)}$

$= \frac{100 + 9 + 8 + 9}{100 + 9 + 8 + 9 + 100 + 1 + 2 + 1}$

$= \frac{126}{230} = 0.548$

$= 0.548$

3) Micro average f_1-score

$$= \frac{TP(A)+TP(B)+TP(C)+TP(D)}{TP(A)+TP(B)+TP(C)+TP(D)+\frac{1}{2}\left(FP(A)+FP(B)+FP(C)+FP(D)+FN(A)+FN(B)+FN(C)+FN(D)\right)}$$

$$= \frac{100+9+8+9}{100+9+8+9+\frac{1}{2}(0+8+2+10+12+100+1+2+1)}$$

$$= \frac{126}{126+\frac{1}{2}(208)}$$

$$= \frac{126}{126+104} = \frac{126}{290} = 0.548.$$