Unit - 3

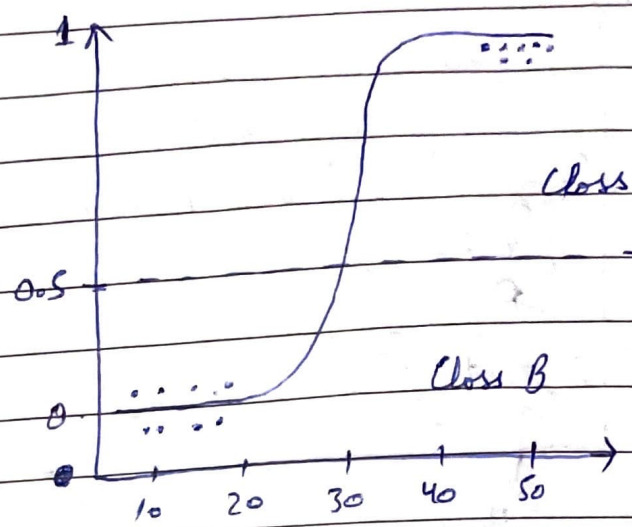
Logistic Regression :- It is used for predicting the categorical dependent variable using a given set of independent variables. It predicts the output of a categorical dependent variable but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

$$Y = \frac{1}{1+e^{-x}}$$ → Sigmoid function.

$e = 2.718$ (eulers constant)

Categorical variable = Yes/No, 0/1, True/False.
Linear Reg is used for solving Reg. problems whereas Log Reg. is used for solving the classification problem.

Eg :- Email Detection
Spam ; No - Spam in gmail



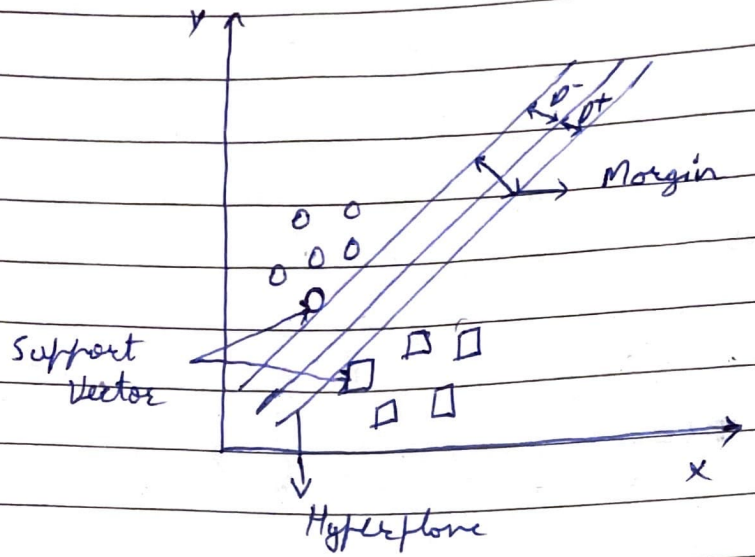
- Class A data has full posibility of occurence
- Class B data has no posibility of occurence
- Data lying on 0.5 is a rare situation and it is unclassify data (iska koi ni hoga bekar h)

Binomial Multinomial Ordinal

Binomial :- 0, 1 Pass, Fail (2 possible value)
Multinomial :- Cat, Dog, Sheep (3 or more —)
Ordinal :- Low, High, Medium (3 or more —)

⇒ Support Vector Machine (SVM) :- It is used for classification as well as regression problem. The goal is to create the best line or decision boundary that can segregate n-dimensional space into classes. This best decision boundary is known as hyperplane. The closest point to the hyperplane are known as support vector. Eg :- Face detection, image classification, text categorization etc.

Linear Non-linear



- Sum of D^- and D^+ is Margin.
- Closest point to hyperplane is support vector.
- Bez of support vector and hyperplane the data is seperated in two diff classes.

Linear SVM :- It is used for linearly separable data, which means if a dataset can be classified into two classed by using a single straight line.

Non-Linear SVM :- cannot be (used for non-linear data)

⇒ Naive Bayes :- It is a probalistic classifier which means it predicts on the basis of the probability of an object. Eg :- Spam filteration, Sentimental analysis and classifying articles.

Bayes Theorem :- $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$

Eg :- If we are searching for a fruit which has attributes like -

Page :
Date: / /

Q Fruit = {Yellow, Sweet, Long}

Fruit	Yellow	Sweet	Long	Total (Fruits)
Orange	350	450	0	650
Banana	400	300	350	400
Others	50	100	50	150
Total	800	850	400	1200

Ans $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$

Orange (Yellow) :-

$= P(\text{Orange}|\text{Yellow}) \times P(\text{Yellow})$

$P(\text{Orange})$

$= \frac{\frac{350}{800} \times \frac{800}{1200}}{\frac{650}{1200}} = 0.53\%$

$P(\text{Sweet}|\text{Orange}) = 0.69\%$

$P(\text{Long}|\text{Orange}) = 0\%$

$P(\text{Fruit}|\text{Orange}) = 0.53 \times 0.69 \times 0 = 0$

$P(\text{Fruit}|\text{Banana}) = 1 \times 0.75 \times 0.87 = 0.65$

$P(\text{Fruit}|\text{Others}) = 0.33 \times 0.66 \times 0.33 = 0.072$

Banana has the highest probability.

Output → Fruit = Banana

Gaussian — Multinomial — Bernoulli

Page :
Date: / /

Bernoulli :- It works similar to Multinomial classifier but the predictor variables are the independent ~~boolean~~ booleans variables. Such as if a particular word is present or not in a document. (It is based on binary nature as it tells something is true or false.

$P(\text{Success}) = p$

$P(\text{Failure}) = q = 1-p$

X = Random Variable

$X = 1$

$X = 0$

$$P(X) = \begin{cases} p & (\text{If } X=1) \\ q & (\text{If } X=0) \end{cases}$$

Multinomial :- This is mostly used for document classification problem ie whether a document belongs to a category of sports, politics etc. The features used by the classifier are the frequency of the words present in a document (Used for finding occurence of a word (count))

Gaussian :- When the predictors take up a continuous value and are not discrete, we assume that these values are sampled from a gaussian distribution. (Used for continuous values)

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(\frac{-(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$

(Definition is enough for all types)

Page :
Date: / /

⇒ K-Nearest Neighbor (KNN) : It assumes the similarity b/w the new data and available data and put the new data into the category that is most similar to available categories. It is also called lazy learner Algo bcz it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs action on dataset. Eg : query $x = (Maths = 6, CS = 8)$, $K = 3$

	Maths	CS	Result
1)	4	3	Fail
2)	6	7	Pass
3)	7	8	Pass
4)	5	5	Fail
5)	8	8	Pass

We have to tell whether x is pass or fail

Euclidean distance : $d = \sqrt{(x_0 - x_A)^2 + (x_0 - x_A)^2}$

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

✓ II $\sqrt{(6-6)^2 + (8-7)^2} = 1$

✓ III $\sqrt{(6-7)^2 + (8-8)^2} = 1$

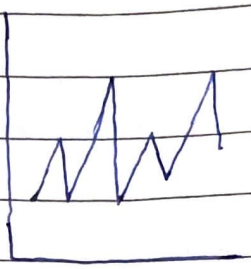
IV $\sqrt{(6-5)^2 + (8-5)^2} = 3.16$

✓ V $\sqrt{(6-8)^2 + (8-8)^2} = 2$

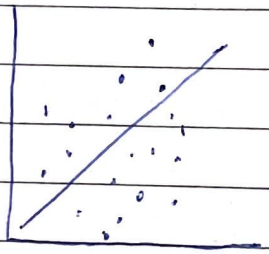
II, III, V are nearest (lowest) to x which is Pass = 3, Fail = 0

So we declare x as pass. (KNN assumes it as pass and put it in data)

⇒ Overfitting :- It occurs when model tries to cover all the data points or more than the required data points present in the dataset. Bcz of this the model starts caching noise and inaccurate value present in the dataset. It has LB and HV.



Underfitting :- It occurs when our machine learning model is not able to capture the underlying trend of the data. (Model is not able to learn enough).



⇒ Regularization :- It is a technique to prevent the model from overfitting by adding extra features to it. It will allow to maintain all variables or features in the model by reducing the magnitude (value) of the variables. Hence it maintains accuracy as well as generalization of the model. (It shrinks the value of variables to zero which are not that much important.) or regularize them

Page :
Date: / /

Techniques :- Ridge Reg.
Losso Reg.

Ridge Reg :- The cost function is altered by adding the penalty term to it. We can calculate it by multiplying with the lambda to the squared weight of each individual feature. It is also called L2 regularization.

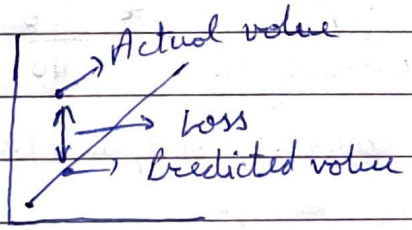
$$R = \text{Loss} + \lambda \|W\|^2$$
(Penalty) (Lambda)

As the value of λ increases the magnitude (value) of co-efficient decreases or shrinks to zero but not exactly zero (greater than 0).

$$\|W\|^2 = W_1^2 + W_2^2 + W_3^2 \text{ ------ } W_n^2$$

Least Absolute and Selection Operator (LASSO) :- It is similar to Ridge expect that the penalty term contains only absolute weight instead of a square of weights. It can shrink a value to exact 0. It is also known as L1 reg.

$$\text{LOSSO } R = \text{Loss} + \lambda \|W\|$$



⇒ Confusion Matrix :- It is used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. Also known as error matrix.

Page :
Date: / /

It evaluates the ~~model~~ performance of classification model, when they make prediction on test data and tells how good our model is. It also tells the error made by the classifier. Eg :-

165	No	Yes	
	50	10	
No	TN	FP	60
	5	100	
Yes	FN	TP	105
	55	110	

FP = Type-I error
FN = —— II ——

Calculations :-

- Accuracy = $\frac{TP + TN}{Total} = \frac{100+50}{165} = 0.91$

It defines how often the model predicts the correct output.

- Error Rate :- 1 - accuracy or $\frac{FP+FN}{Total} = 0.09$

It defines how often the model gives the wrong prediction

- Precision :- $\frac{TP}{\text{Predicted Yes}} = \frac{100}{110} = 0.64$

Number of correct output provided by the model (Eg)

- Recall :- $\frac{TP}{\text{Actual Yes}} = \frac{100}{105} = 0.95$

Out of total positive classes, how often our model predicted correctly. Eg (5/1) (Recall is low) (4 still remain)

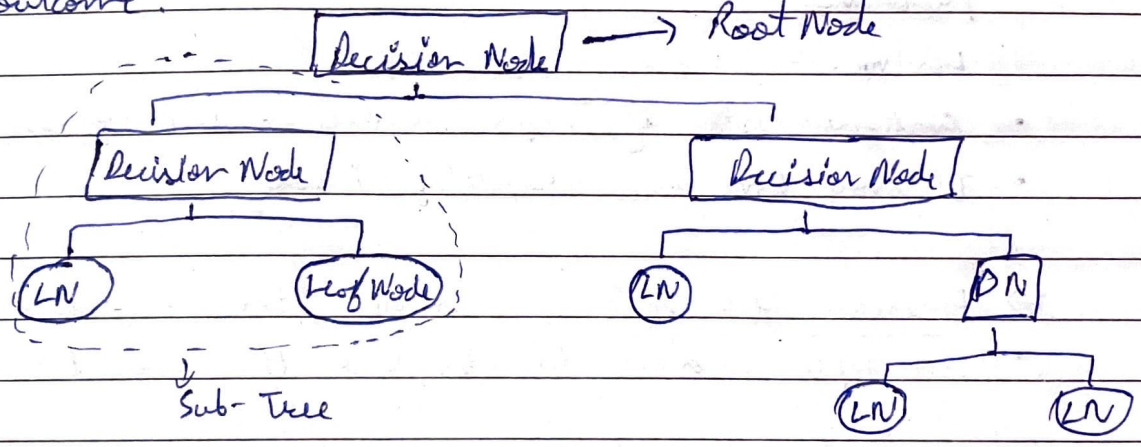
- F1-score :- $2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}$

It helps to evaluate the recall and precision at same time.

- Specifity : $\frac{TN}{TN+FP} = \frac{50}{50+10} = 0.83$

The true negative that are correctly predicted by the model.

⇒ Decision Tree : It is a graphical representation for getting all the possible solutions to a problem/ decision based on given conditions. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represent the outcome.



Information Gain : It is the measurement of changes in entropy after the segmentation of a dataset based on an attribute.

Entropy : It is a metric to measure the impurity in a given attribute. It specifies randomness in data.

Gini Index :- It is a measure of impurity or purity used while creating a decision tree in the CART (Classification and Regression Tree) algorithm.

Pruning :- It is a process of deleting the unnecessary nodes from a tree in order to get optimal decision tree.

Cost Complexity | Reduced Error

Advantages :-

- Simple to understand. Can be very useful for solving decision-related problem.
- It helps to think about all possible outcome for a problem.

Disadvantages :-

- contains lots of layer, which makes it complex.
- It may have an overfitting issue.

⇒ Ensemble learning :- It combine the decisions from multiple models to improve the overall performance.

Random Forest :- It is a classifier that contain a number of decision Tree on various subsets of the given dataset and takes the average to improve the predictive accuracy of the dataset.

- It takes less training time
- It predicts output with high accuracy
- It can also maintain accuracy when a large proportion of data is missing.