Machine Learning

DOMS Page No.
Date 08 / 01 / 2023

1) What is Machine Learning ?

↳ ML is a subfield of artificial ~~integ~~ intelligence, which is broadly defined as the capability of machine to imitate ~~integr~~ intelligent human behaviour.

* Biggest Confusion : AI Vs ML Vs Deep Learning

1) AI :- A technique which enables machines to mimic human behaviour.

2) ML :- Subset of AI technique which use statistical methods to improve with experience.

3) Deep Learning :- Subset of ML which make the computation of multi-layer neural network feasible.

* Types of ML -

1) Supervised learning

↳ Supervised Learning is defined by its use of labelled datasets to train algorithms that to classify data or predict outcomes accurately.
↳ Basically, when the o/p is available then it is known as supervised learning. (X → Y)

* Supervised Learning Algo

↳ Linear Regression, Random Forest, Support Vector machines (SVM)

DOMS Page No.
Date / /

Un
2) ~~Supervised~~ Learning

↳ Unsupervised learning is used to analyze and cluster unlabeled datasets.
↳ These algorithms discover hidden patterns or data grouping.
↳ When there is no o/p available only i/p is available is known as unsupervised learning. (X not Y)

* Algorithms

↳ K-means, Hierarchical clustering

3) Reinforcement Learning

↳ It is a type of ML method where an intelligent agent interacts with the environment and learns to act within that.

* Supervised Learning Algorithms

1) Linear Regression

↳ What is Regression?

↳ Regression analysis is a form of predictive modelling technique which investigate the relationship B/W a dependent and independent variable.

* Uses of Regression

↳ Determining the strength of predictors
↳ Forecasting an effect , and
↳ Trend forecasting

* Linear vs logistic Regression

1) Linear Regression

↳ The data is modelled using a straight line.
↳ It is used with continues variable.
↳ It gives the value of the variable as o/p.
↳ It's accuracy is measured by loss , R squared etc.

2) Logistic Regression

↳ The probability of some obtained event is represented by as a linear funⁿ combination of predictor value
↳ It is used with categorical value.
↳ It gives probability of accurance of event.
↳ It's accuracy is measured with precision, Recall, F1 score , ROC curve , confusion matrix. etc.

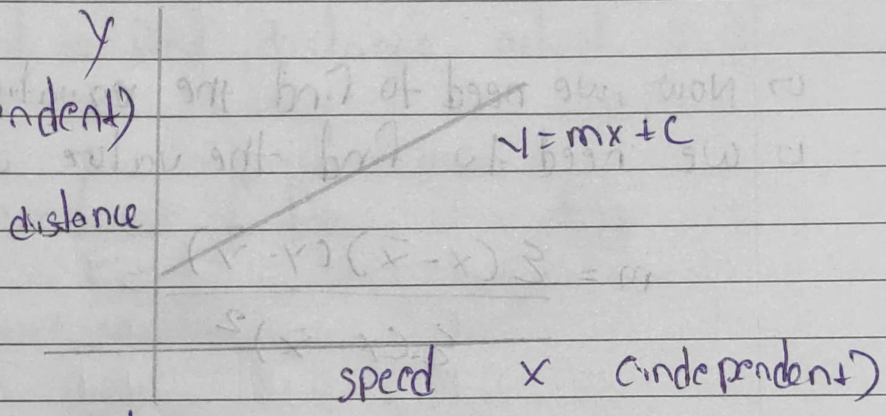
* Selection Criteria - LR

↳ Classification and Regression Capabilities
↳ Data Quality
↳ computational complexity

* **where is LR used ?**

↳ Evaluating Trends and Sales Estimates.
↳ Analyzing the impact of Price changes.
↳ Assessment of Risk in financial service and insurance domain.

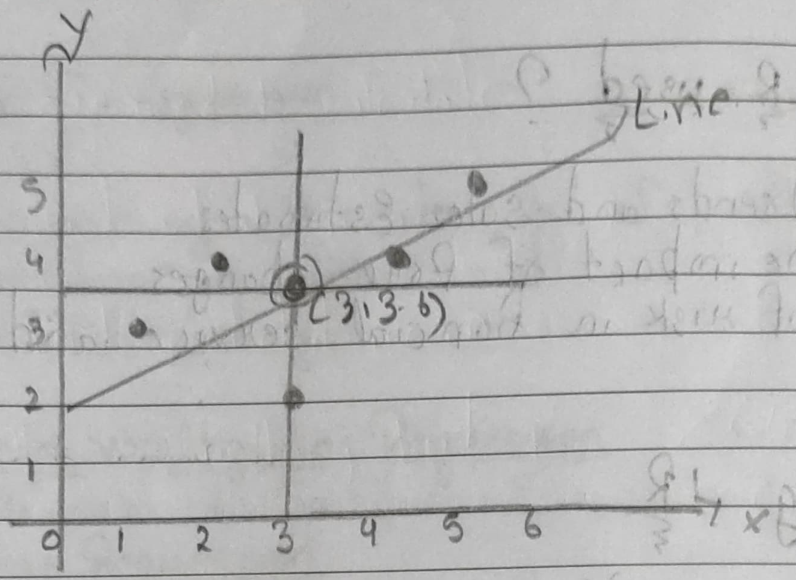
* **Understanding LR**



y = distance travelled in a fixed duration of time.
x = speed of vehicle
m = +ve slope of the line
c = y intercept of the line.

* **Mathematical implementation.**

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	3	-2	-0.6	4	1.2
2	4	-1	0.4	1	-0.4
3	2	0	-1.6	0	0
4	4	1	0.4	1	0.4
5	5	2	1.4	4	2.8
$\bar{x}=3$	$\bar{y}=3.6$			$\Sigma=10$	$\Sigma=4$



↳ Now, we need to find the equation of Regression line

↳ We need to find the value of m & c.

$$m = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$= 4/10 = 0.4$$

$$\boxed{m = 0.4}$$

Now,

$y = mx + c$

$3.6 = 0.4 \times 3 + c$

$3.6 = 1.2 + c$

$3.6 - 1.2 = c$

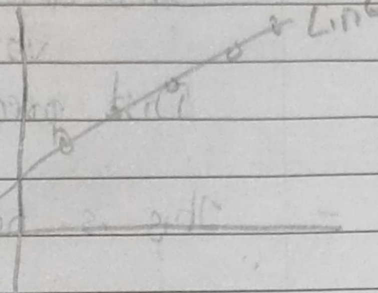
$c = 2.4$

$$\boxed{c = 2.4}$$

dist = actual - mean

For given $m = 0.4$ & $c = 2.4$, let's predict values for y for $x = \{1, 2, 3, 4, 5\}$, $y = 0.4x + 2.4$

$y = 0.4 \times 1 + 2.4 = 2.8$
$y = 0.4 \times 2 + 2.4 = 3.2$
$y = 0.4 \times 3 + 2.4 = 3.6$
$y = 0.4 \times 4 + 2.4 = 4.0$
$y = 0.4 \times 5 + 2.4 = 4.4$



Now, we need to find distance actual & predicted value.

* **Lets check the goodness of fit**

↳ Well use R square method.

1) What is R square?
↳ R squared value is a statistical measure of how close the data are to the fitted regression line.
↳ It is also known as coefficient of determination, or the coefficient of multiple determination.

* **Calculation of R^2 method.**

x	yp
1	2.8
2	3.2
3	3.6
4	4.0
5	4.4

→ Now we need to calculate,

Dist actual - mean

vs

Dist predicted - mean

This is nothing but $R^2 = \frac{\Sigma (yp - \bar{y})^2}{\Sigma (y - \bar{y})^2}$

x	y	$y-\bar{y}$	$(y-\bar{y})^2$	yp	$(yp-\bar{y})$	$(yp-\bar{y})^2$
1	3	-0.6	0.36	2.8	-0.8	0.64
2	4	0.4	0.16	3.2	-0.4	0.16
3	2	-1.6	2.56	3.6	0	0
4	4	0.4	0.16	4.0	0.4	0.16
5	5	1.4	1.96	4.4	0.8	0.64
	3.6		5.2			1.6

$R^2 = \frac{1.6}{5.2}$

$R^2 \approx 0.3$ (which is not best fit points are far away)

when $R^2 \approx 1$, it will best fit.

2) Logistic Regression (Classification Algorithm)

↳ Logistic Regression produces results in a binary format which is used to predict the outcome of a categorical dependent variable.

↳ So, the outcome should be discrete / categorical such as:

=> 0 OR 1

=> Yes OR NO

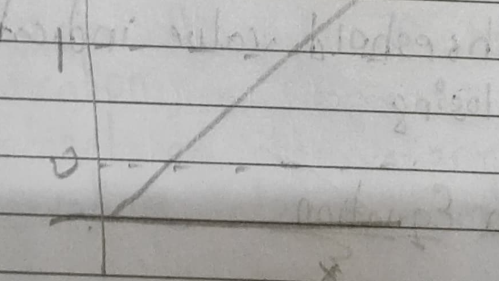
=> True OR FALSE

=> High OR LOW

* why not Linear Regression ?

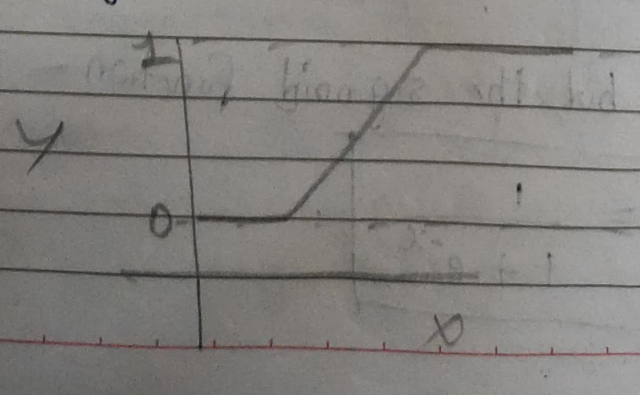
↳ In linear Regression the o/p has a specific range.

↳ But, in Logistic Regression the o/p has only 2 values it can be either 0 or 1.



we dont need values below 0 and after 1.

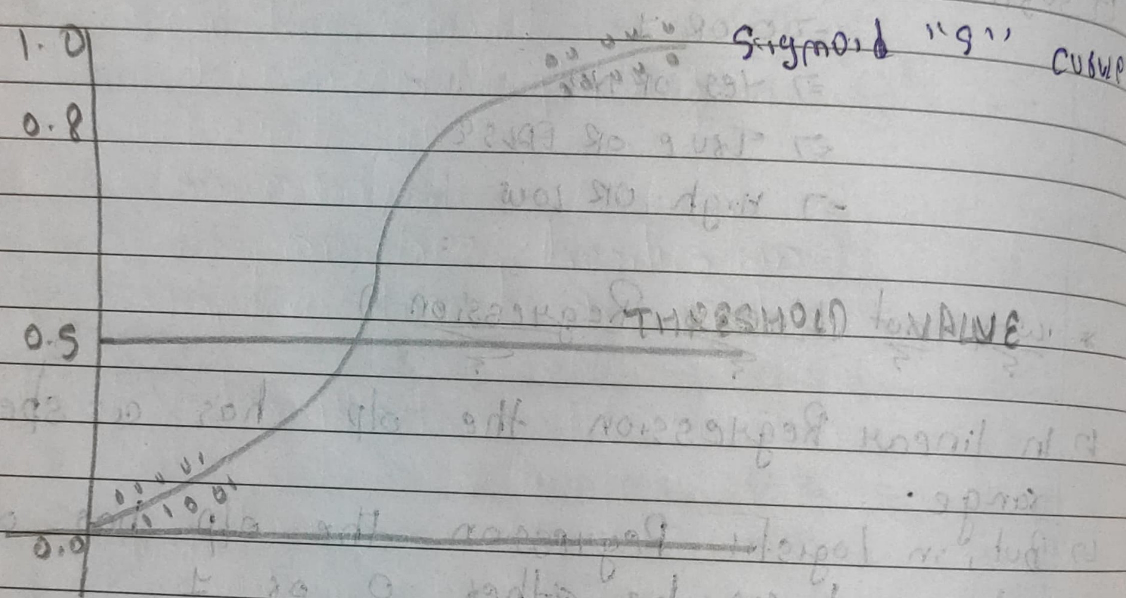
↳ So, the graph should like -



↳ Here we'll widely used sigmoid function wh
converts -infinity to infinity value to discrete valu
(in the form of 0 or 1).

* **Logistic Regression Curve**

↳



↳ with this the threshold value indicates the probabili
of winning or losing

* **Logistic Regression Equation**

↳ $\log\left[\frac{Y}{1-Y}\right] \Rightarrow Y = C + B1X1 + B2X2 + \cdots$

↳ It is nothing but the sigmoid function –

$$\boxed{Y = \frac{1}{1+e^{-x}}}$$

DOMS Page No.
Date / /

* Linear VS Logistic Regression

Linear Regression	Logistic Regression
1) Continuous variables	1) Categorical variables
2) Solve Regression problems	2) solve classification Problems
3) Straight Line Graph.	3) S curve.
4) $Y = mx + c$	4) $Y = \frac{1}{1+e^{-x}}$

* Classification Algorithm

- The classification algorithm is a supervised Learning technique that is used to identify the category of new observations on the basis of training data.
- In classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups.
- Such as, Yes or No, 0 or 1, Spam or Not spam, cat or Dog etc.
- Unlike regression, the o/p variable of classification is a category, not a value such as Green or Blue etc.
- Hence Classification Algorithm is a Supervised learning technique

Date / /

Page No.

↳ Classification Algorithms can be further divided into mainly two category :-

·) Linear Models

↳ Logistic Regression

↳ Support vector Machines (SVM)

·) Non-Linear Models

↳ K-Nearest Neighbours

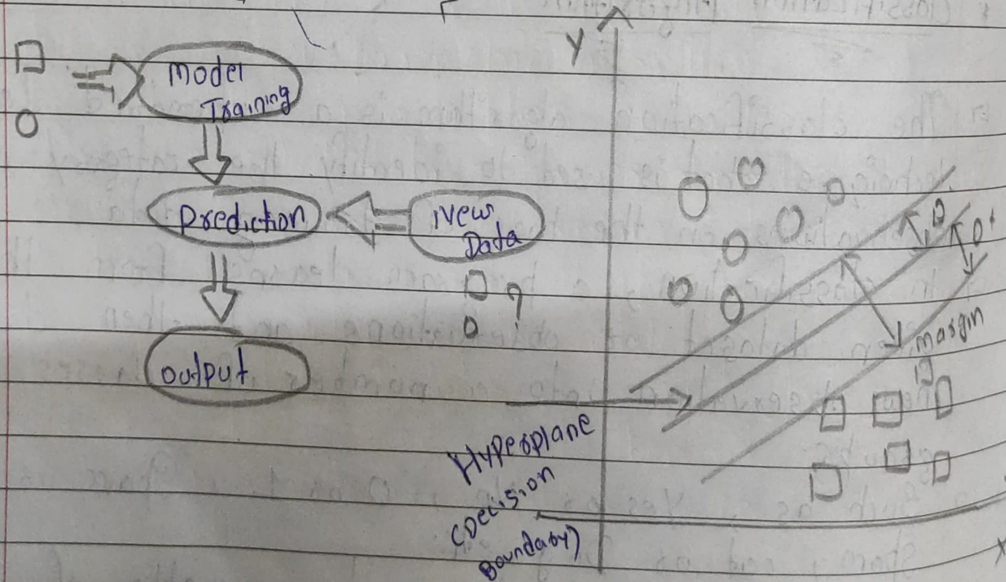
↳ Kernel SVM

↳ Naive Bayes

↳ Decision Tree Classification

↳ Random Forest Classification.

③ Support Vector Machines (SVM)



↳ Margin = D⁻ + D+

↳ There is a big role of Margin in ~~selecting the~~ which hyperplane exists.

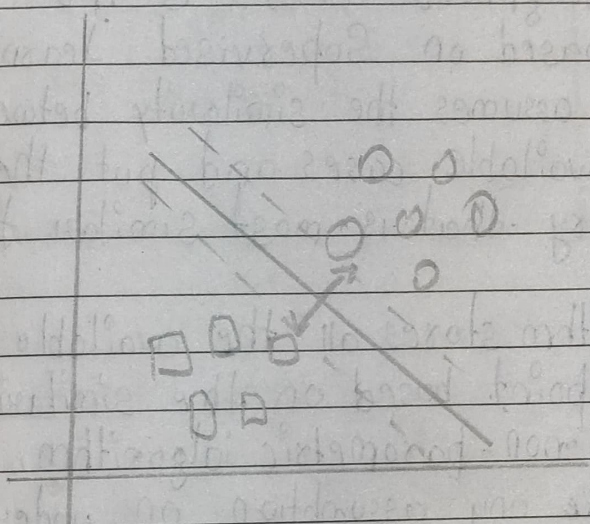
DOMS Page No.
Date / /

→ What is support vectors?

→ Points which are closer to opponents class are knowns as support vectors.

* Maximal Margin Hyperplane

→



→ The margin should be of maximum width. which Hyperplane margin has maximum width will choose.

* Theory

→ SVM is one of the most popular supervised Learning algorithms, which is used for classification problems.

→ The goal of the SVM algo is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can put easily new datapoint in the correct category in the future.

→ The best decision boundary is known as Hyperplane.

↳ SVM chooses the extreme points / vectors that help in creating the hyperplane.
↳ These extreme cases are called as support vectors, and hence algo is termed as SVM.

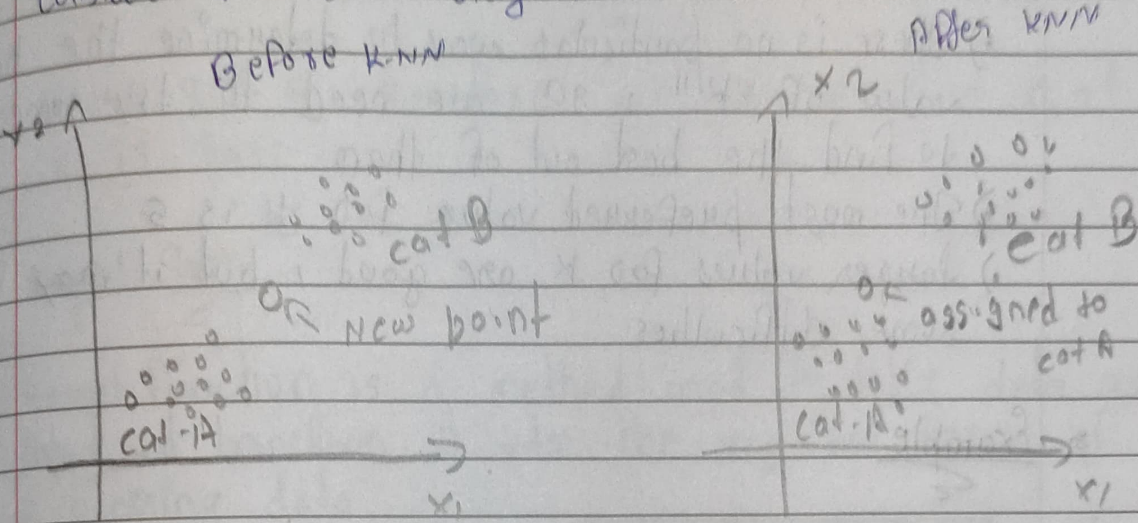
(9) K-Nearest Neighbour (KNN)

↳ K-Nearest Neighbour (KNN) is the simpliest ML algorithms based on Supervised learning.
↳ K-NN algo assumes the similarity between the new case / data and available cases and put the new cases into the category that is most similar to the available categories.
↳ K-NN algorithm stores all the available data and classifies a new data point based on the similarity.
↳ KNN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
↳ It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification it performs an action on the dataset.
↳ KNN algo at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to that data.

* Why do we need KNN algorithm?

↳ Suppose there are two categories, i.e. category A and B, and we have a new datapoint x1, so this point lies on which category.

- To ~~stor~~ solve this problem, we need a KNN algo.
- With the help of KNN, we can easily identify the category or class of a particular dataset.
- Consider the below diagram :-



* How does K-NN work?

- The KNN works on the below algorithm -

Step 1 :- Select the number K of the neighbors

Step 2 :- Calculate the Euclidean distance of K no of neighbours.

Step 3 :- Take the K nearest neighbors as per the Euclidean distance.

Step 4 :- Among these K neighbors, count the no. of data points in each category

Step 5 :- Assign new data points to that category for which the no. of neighbor is max.

Step 6 :- our model is ready.

* How to select the value of K in K-NN?

↳ Below are some points to remember while selecting the value of K in KNN –

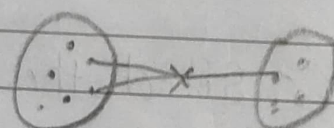
- These is no particular way to determine the best value of 'K', so we need to try some values to find the best out of them.
- The most preferred value for K is 5.
- Larger values for K are good, but it may find some difficulties.

* Example

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

Euclidien-distance :-

$$d = \sqrt{|x_{01} - x_{A1}|^2 + |x_{02} - x_{A2}|^2}$$



query :- x = (maths = 6, CS = 8), K = 3.

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 $\sqrt{|6-8|^2 + |8-8|^2} = 2$

closest Neighbour

2) P , 3) P , 5) P = 3 Pass
0 fail

$3 > 0$

↳ Here $3 > 0$, therefore we can predict that x is Pass.

* **Kernel SVM**

↳ Kernel function is a method used to take data as input and transform it into the required form of processing data.

↳ Basically it convert n dimensional space into n+1 dimensional space

* **SVM imp Equations** find ω, b ?

↳ N points $(x_1, y_1)(x_2, y_2) \cdots (x_n, y_n)$

↳ $y_i = +1 / -1$

↳ $\min \frac{1}{2} \|w\|^2$ Subject to $y_i(wx_i - b) \geq 1$

↳ $\omega = \sum_{i=1}^{n} \alpha_i y_i \bar{x}_i$ where $\alpha = (\alpha_1, \alpha_2 \ldots \alpha_n)$ (vector)

↳ Equation for $\alpha \Rightarrow \phi(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1} \alpha_i \alpha_j y_i y_j$

$(\bar{x}_i - \bar{x}_j)$ S.t $\sum_{i=1}^{N} \alpha_i y_i = 0$,

$\alpha_i \geq 0$ for

$$b = \frac{1}{2}\left(\left(\min_{i=y_i+1}(\bar{w}\bar{x_i}) + \max_{i-y_i=-1}(\bar{w}\bar{x_i})\right)\right)$$

unit 3.

① Confusion Matrix

↳ A confusion Matrix is used for evaluating the performance of a classification model.

		Actual +ve	Actual -ve
predicted	+ve	TP	FP
	-ve	FN	TN

Ex // Confusion matrix of Disease detection.

n = 200	Predicted NO	Predicted Yes	
Actual NO	TN = 100	FP = 15	115
Actual Yes	FN = 5	TP = 80	85
	105	95	

TN :- Patient not have disease & model also says same.
TP :- Patient have disease & model also says same.
FP :- Patient actually not have disease, but model says they have.
FN :- Patient have disease but model says they dont.

* Basic Calculations on the basis of Confusion Matrix.

1) Accuracy :- No. of predictions that are right.
↳ Accuracy = (TP + TN) / Total.

2) Error rate :- No. of Predictions that are wrong.
↳ Error rate = (FP + FN) / Total.

3) Precision :- When the model predicts the +ve, how often is ...

↳ Precision = TP / (TP + FP)

4) Recall :- When it actually yes, how often does it pred... yes.

↳ Recall = TP / (TP + FN)

In previous example,

1) Accuracy = (TP + TN) / N = 180 / 200
$= \boxed{0.9}$

2) Error rate = (FP + FN) / N = 20 / 200
$= \boxed{0.1}$

3) Precision = TP / (TP + FP) = 80 / 95
$\boxed{= 0.86}$

4) Recall = TP / (TP + FN) = 80 / 85
$\boxed{= 0.94}$

5) F1 score = $2 \cdot \dfrac{\text{Precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

$= 2 \times \dfrac{0.86 \times 0.94}{0.86 + 0.94}$

$= 2 \times \dfrac{0.8084}{1.8}$ $= 2 \times 0.449$

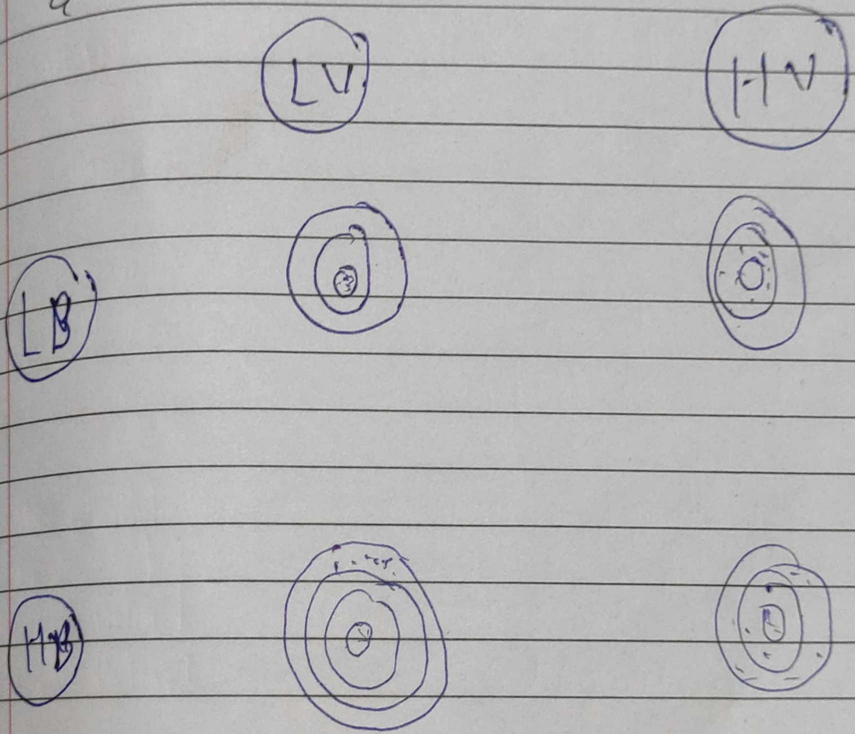
$\boxed{= 0.89}$

2) Validation :-

-> Types :-

1) Leave One Out Cross validation.
2) K-fold cross validation.
3) Hold out cross validation.
4) validation Set Approach.

3) Bias & Variance

-> 

-> Bias :- Gap between Actual and Predicted value.

-> High Bias :- Predicted values are far away from actual values.

-> low Bias :- Predicted value are close from actual values.

DOMS Page No.
Date / /

↳ Variance :- How much scattered predicted value

Note : There should be low bias & High variance.

(4) Bias Variance Tradeoff.

↳