Omkar Gurav
BE IT
8048

MLA Theory Assignment 3

Q.1. Consider following dataset & predict the class of new instance X using Naive Bayes Classification algorithm.

Tid	Refund	Marital status	Taxable Amount	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

X = (Refund = No, Marital status = Married, Income = 120k)

→ ① Learning phase for Refund

Refund	Evade = Yes	Evade = No
Yes	0/3	3/7
No	3/3	4/7

② Learning phase for Marital status

Marital status	Evade = Yes	Evade = No
Single	2/3	2/7
Married	0/3	4/7
Divorced	1/3	1/7

classmate

Date

Page

③ Learning phase for Taxable Amount

if ~~class~~ Evade = No, Sample mean = $\frac{125+100+70+120+60+220+75}{7}$

$= \frac{770}{7}$

Sample mean = 110

Sample variance = $\frac{(15)^2+(10)^2+(40)^2+(10)^2+(-50)^2+(110)^2+(-35)^2}{6}$

$= \frac{17850}{6}$

Sample variance = 2975

If Evade ~~class~~ = Yes, Sample mean = $\frac{95+85+90}{3} = \frac{270}{3} = 90$

sample variance = $\frac{(5)^2+(-5)^2+0^2}{2} = \frac{50}{2} = 25$

For new instance X = (Refund = No, Marital status = Married, Income = 120k)

P(Refund = No | Evade = Yes) = 1

P(Marital status = Married | Evade = Yes) = 0

P(Income = 120k | Evade = Yes) = $\frac{1}{\sqrt{2\pi}\,(54.54)} e^{(120-90)^2/2(25)}$

$= 0.0073 \times e^{18}$

P(Refund = No | Evade = No) = 4/7

P(Marital status = Married | Evade = No) = 4/7

P(Income = 120k | Evade = No) = $\frac{1}{\sqrt{2\pi}\,(54.54)} e^{(120-110)^2/2(2975)}$

$\simeq 0.0072$

$P(Yes|x) = P(Refund = \cancel{Yes}\ No \mid Yes) \times P(Marital\ status = Married \mid Yes) \times$
$P(Income = 120k \mid Yes)$
$= 1 \times 0 \times 0.0073 \times e^{18}$

$P(Yes|x) = 0$

$P(No|x) = P(Refund = No \mid No) \times P(Marital\ status = Married \mid No) \times$
$P(Income = 120k \mid No)$
$= 4/7 \times 4/7 \times 0.0072$

$P(No|x) = 0.0023$

$P(No|x) > P(Yes|x)$

Thus, the label for x will be "No".

Q.2. Explain Expectation-maximization algorithm.

→

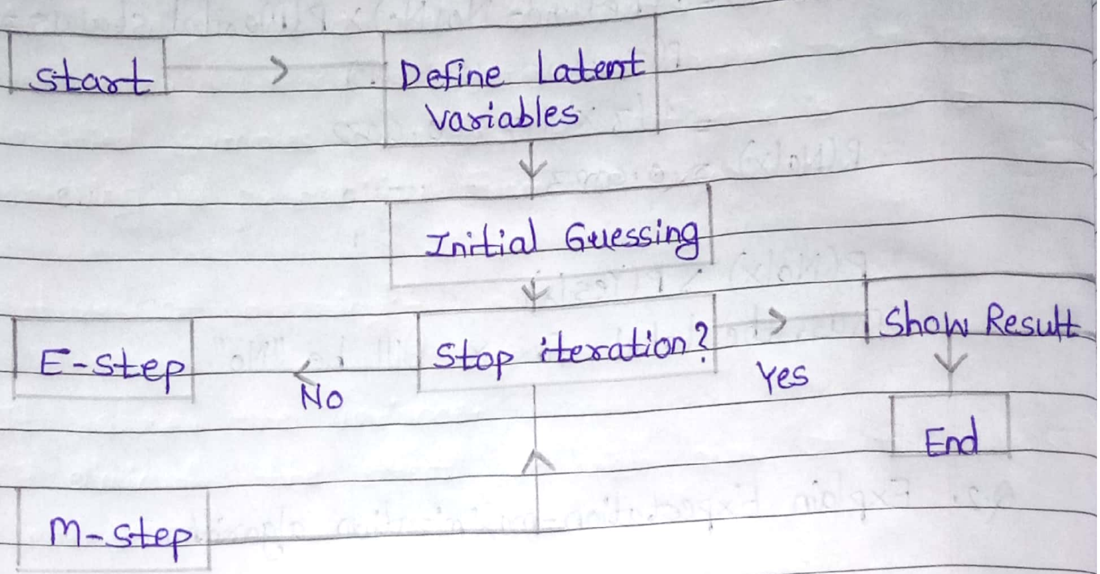
The EM algorithm is the technique that can be deployed in order to determine the local maximum likelihood estimates or maximum a posterion estimates for latent variables in statistical models.

Being an iterative approach, the EM algorithm revolves amid 2 modes, the first mode estimates the missing or latent variables, called E-step & the second step optimizes the parameters of the model that explains data more clearly, called M-step. i.e.

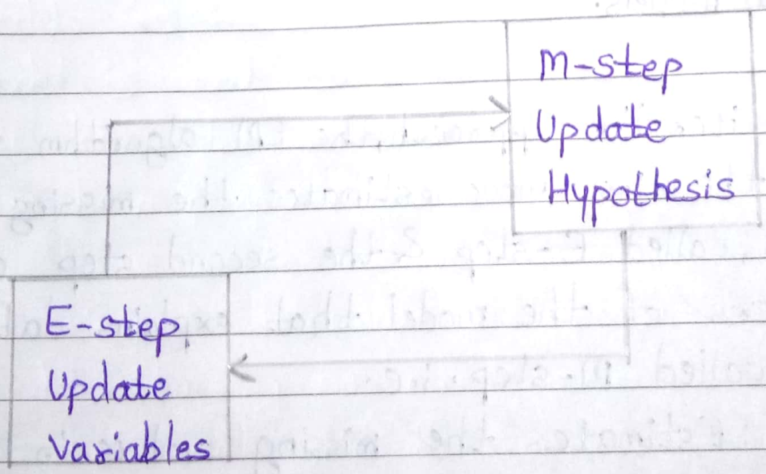
E-step : Estimates the missing values in the dataset.

M-step : Maximize the model parameters while the data is present.

The algorithm is used for predicting these values or in computing missing or incomplete data, given the generalised form of probability distribution that is connected with these latent variables.



Workflow of EM algorithm

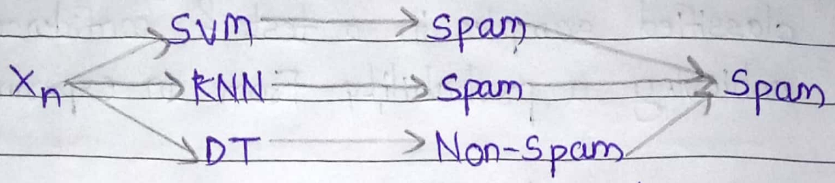


Working of E-step & M-step

Q.3. What is ensemble learning ? Explain bagging & boosting, stacking in brief.

→

Ensemble classifier approaches are as follows:
1. To generate a multiple classifiers
2. Each voting based on test instance like shown below:



Ensemble classifier approach

3. Consider majority as classification

Ensemble methods vary from each other. They have different training training strategy & combination method.

Bagging stands for Bootstrap Aggregation. It works by training multiple models on different samples & took the average of their prediction.
It works in underfitting, overfitting.
Analysis of Bagging: Expected error = Bias + Variance

Main features of bagging:-

① It decreases variance of the base model with the hel of averaging without any change in the bias.
② It is useful in models with high variance & noisy da
③ It is useful in overfitted base model as it has a high dependency on training data.

④ Accuracy is achieved in a model with the help of using its multiple copies.

Boosting works on PAC i.e. Probably Approximately Correct Framework.
PAC learning has 2 main components: accuracy & confidence. Accuracy is percentage drawn from correctly classified samples in a test & confidence is about achieving a probability from an experiment.

Features of boosting:-

① It reduces the variance.
② It eliminates the effect of high bias of the weak learner
③ Train versus test errors, performance is:
Train errors can be driven close to 0
Test errors do not reflect over-fitting

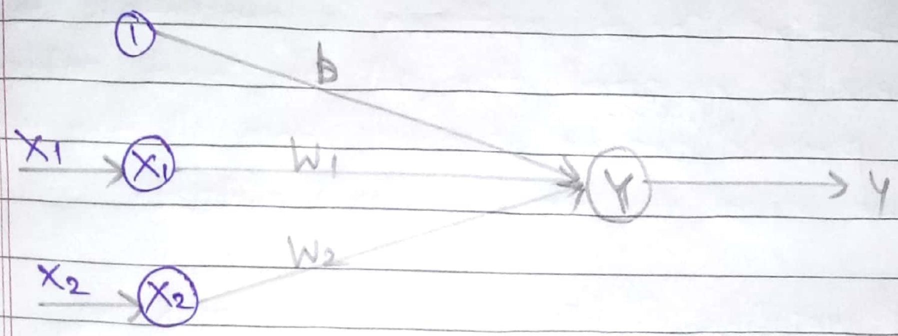
Stacking is the art of using predictions from multiple models like 'Features' & train a new model; this new model is used to make predictions on test data. It looks for whether a Learning of training data has been properly learned.

Stacking & votes are 2 approaches for combining models. Stacking has no learning at the second tier or meta-level while it combines classifiers by a voting scheme.
A label which is most often allocated to a certain instance is Preferred as the correct Prediction when using voting.

Q.4. Implement AND Function using perceptron network using following bipolar inputs & target.

X1	X2	T
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1

→



Perceptron network for AND function

The input patterns are presented to the network one by one. When all the four input patterns are presented, then 1 epoch is said to be completed. The initial weights & threshold are set to 0 i.e. $w_1 = w_2 = b = 0$ & $\theta = 0$. The learning rate α is set to 1

For 1st input pattern, $x_1 = 1$, $x_2 = 1$ & $t = 1$, with weights & bias, $w_1 = 0$, $w_2 = 0$ & $b = 0$

Net input $y_{in} = b + x_1 w_1 + x_2 w_2 = 0 + 1 \times 0 + 1 \times 0 = 0$
The output y is calculated by applying activations over the net input :

$$y = f(y_{in}) = \begin{cases} 1 & \text{if } y_{in} > 0 \\ 0 & \text{if } y_{in} = 0 \\ -1 & \text{if } y_{in} < 0 \end{cases}$$

Here, we have taken $\theta = 0$. Hence, when $y_{in} = 0$, $y = 0$.
But here $t = 1$ & $y = 0$ so $t \neq y$ hence weight updation takes place.

$w_i(new) = w_i(old) + \alpha t x_i$

$w_1(new) = w_1(old) + \alpha t x_1 = 0 + 1 \times 1 \times 1 = 1$

$w_2(new) = w_2(new) + \alpha t x_2 = 0 + 1 \times 1 \times 1 = 1$

$b(new) = b(old) + \alpha t = 0 + 1 \times 1 = 1$

The weights $w_1 = 1, w_2 = 1, b = 1$ are the final weights after first input pattern is presented. The same process is repeated for all the input patterns. The process can be stopped when all the targets become equal to the standard output.

Final weights & bias after 2^{nd} epoch are
$w_1 = 1, w_2 = 1, b = -1$
Since the threshold for problem is 0, eq^n of separating line is $x_2 = -\frac{w_1}{w_2} x_1 - \frac{b}{w_2}$

Thus, using final weights we obtain

$$x_2 = -\frac{1}{1} x_1 - \frac{(-1)}{1}$$

$$x_2 = -x_1 + 1$$

It can be easily found that the above straight line separates the positive & negative response region.