DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY.
ONLINE SESSIONAL EXAMINATION

B.Tech (CE) sem → 7
subject → Machine Learning

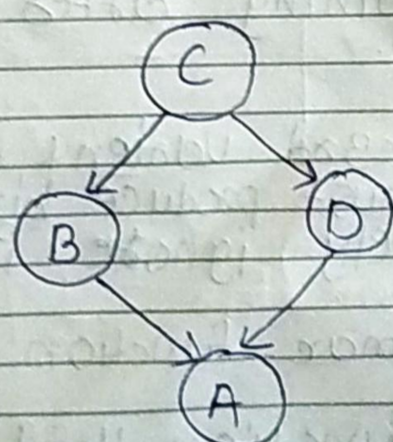
Roll no. → 142
Signature → [signature]

Date :→ 26/8/2020
Time → 10:00 to 11:15
Total pages : 13

Q.1

A. True Because Naive Bayes classifier makes assumptions & distribution estimation & with the help of it, it ~~tags~~ tries to learn model.

B.



C. figure 1 :→ A & B are independent
& C is dependent on A&B

figure 2 :→ A is dependent on C
B is dependent on C
A & B are not dependent on each other.
C is independent

D. (I) in case of attribute having multiple splites i.e. multiple values of attribute
we should remove splitting attr. from that list ~~been~~ so that in further iteration it is not chosen repeatedly

(II) in binary ~~tary~~ tree there are only 2 options possible not multiple, so only one option is remaining after choosing splitting attr.

(III) the 3rd varient i.e. "A is descrete & must produce binary tree" we can surely ignore it.

E. softmax function

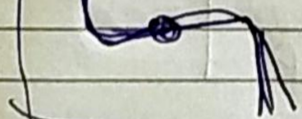
→ a function that takes vector z of k real numbers as input & normalizes it into probability distribution consisting of k probabilities proportional to exponential of the input numbers.

$$\sigma : R^k \to R^k$$

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \quad \text{for } i=1 \text{ to } K$$

F. when gradient descent converges in convex fn, it reaches to minimum value (local) of that point & high probability is there of it being the optimum.

in non convex loss funtion we can't assure the above fact



we might get given point in figure.

Q.3

(A) (I)

Performance matrice

for A,

	Predicted A	Predicted NOT A
actual A	4 TP	4 FN
actual not A	3 FP	5 TN

for B,

	Predicted B	Predicted NOT B
actual B	2 TP	2 FN
actual not B	3 FP	9 TN

for C,

	Predicted C	Predicted not C
actual C	2 TP	2 FN
actual not C	2 FP	10 TN

(1) precision ~~for~~ $= \frac{1}{3}\left(\text{Precision A} + \text{Precision B} + \text{Precision C}\right)$

$$= \frac{1}{3}\left[\frac{TP(A)}{TP(A)+FP(A)} + \frac{TP(B)}{TP(B)+FP(B)} + \frac{TP(C)}{TP(C)+FP(C)}\right)$$

5

$= \frac{1}{3}\left(\frac{4}{4+3} + \frac{2}{2+3} + \frac{2}{2+2}\right)$

$= \frac{1}{3}\left(\frac{4}{7} + \frac{2}{5} + \frac{2}{4}\right)$

$= \frac{1}{3}(0.571428 + 0.4 + 0.5)$

$\doteq 0.49048$

(2) Recall $= \frac{1}{3}(\text{recall}(A) + \text{recall}(B) + \text{recall}(C))$

$= \frac{1}{3}\left(\frac{TP(A)}{TP(A)+FN(A)} + \frac{TP(B)}{TP(B)+FN(B)} + \frac{TP(C)}{TP(C)+FN(C)}\right)$

$= \frac{1}{3}\left(\frac{4}{4+4} + \frac{2}{2+2} + \frac{2}{2+2}\right)$

$= \frac{1}{3}(0.5 + 0.5 + 0.5)$

$= 0.5$

(II) to convert this problem into regression problem,

(B) ~~(I)~~ Logistic ~~equ~~ regression

→ It is is a statistical model that uses logistic function to model a binary dependent variable in its basic form.

→ It is a predictive analysis used to describe data & explain relationship b/w one dependent binary variable & 1 or more nominal, ordinal, interval ~~ratio~~ independent variables.

→ it gives $\boxed{\text{probability}}$ of instance for belonging to some class.

→ uses $\boxed{\text{probabilistic approach}}$

$$h_\theta(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$\therefore h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$$\left(\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \qquad x^T = [1\ x_1 \ldots x_d] \right)$$

$h_\theta(x) =$ estimated $P(y=1 \mid x; \theta)$

$P(y=0 \mid x; \theta) = 1 - P(y=1 \mid x; \theta)$

→ to classify, we have to take some threshold of probability

e.g. for 2 classes. ($y=1$ or 0)

Predict $y=1$ if $h_\theta(x) \geq 0.5$

$y=0$ if $h_\theta(x) < 0.5$

→ for Non-linear boundaries, we can use basis fn expansion to ~~for~~ features same like linear regression.

→ for gradient decent,

→ initialize θ & repeat below until convergence.

$$\theta_0 \leftarrow \theta_0 - \alpha \sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$\theta_j \leftarrow \theta_j - \alpha \left[\sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)}\right) x_j^{(i)} + \lambda \theta_j \right]$$

(simultaneous update for $j = 0 \dots d$)

→ We cant use squared error loss in logistic regression,

because it results in a non-convex optimization

(II) → Prepruning

→ ~~cappos~~ it say that we should stop growing the tree before it reaches the point where it perfectly classifies the training data.

→ halts tree construction early deciding not to split further for given node & it becomes a leaf.

→ for decideing threshold we should take note that

high thresholds → over simplified tree.

low thresholds → very low simplification

⇒ Post pruning

→ removes sub trees from a prepared tree & replace it with a leaf.

→ & that leaf is named the most appropriate class for the subtree which is removed.

→ it results into fast application, easy interpretation, easy implementation but it doesn't effect very high dimensional data much.

→ possibility of dropping important features is there.

Q.2

(A)

C : 1 5 times
    0 5 times

C	1	0
	0.5	0.5

for $C=1$ $x_1=1$ 2 times
            $x_1=0$ 3 times
    $C=0$ $x_1=1$ 3 times
            $x_1=0$ 2 times

x_1 \ C	1	0
1	2/5	3/5
0	3/5	2/5

for $C=1$ $x_2=1$ 2 times
            $x_2=0$ 3 times
    $C=0$ $x_2=1$ 4 times
            $x_2=0$ 1 times

x_2 \ C	1	0
1	2/5	4/5
0	3/5	1/5

$P(C=1 / x_1=1, x_2=1)$

$$= \frac{P(x_1=1, x_2=1 / C=1)\, P(C=1)}{P(x_1=1, x_2=1)}$$

$$= \frac{P(x_1=1/C=1)\, P(x_2=1/C=1)\, P(C=1)}{P(x_1=1, x_2=1, C=0) + P(x_1=1, x_2=1, C=1)}$$

$$= \frac{P(x_1=1/C=1)\, P(x_2=1/C=1)\, P(C=1)}{\left(P(x_1=1/C=0)\, P(x_2=1/C=0)\, P(C=0) + P(x_1=1/C=1)\, P(x_2=1/C=1)\, P(C=1)\right)}$$

$$= \frac{(2/5)(2/5)(0.5)}{(3/5)(4/5)(0.5) + (2/5)(2/5)(0.5)}$$

$$= \frac{(0.4)(0.4)(0.5)}{(0.6)(0.8)(0.5) + (0.4)(0.4)(0.5)}$$

$$= \frac{0.08}{0.24 + 0.08}$$

$$= \boxed{0.25}$$

(B)

$P(D = \text{True} \mid A = \text{False}, C = \text{True}, E = \text{true})$

$$= \frac{P(D=\text{True})\, P(A=\text{False}, C=\text{True}, E=\text{True} \mid D=\text{True})}{P(A=\text{False}, C=\text{True}, E=\text{True})}$$

$P(D = \text{True}) = \cancel{P(D/C)\,P(C)}$

$= P(D=\text{True}, C=\text{True})$
$+ P(D=\text{True}, C=\text{False})$

$= P(D=\text{True}, C=\text{True}, A=\text{True}, B=\text{True})$
$+ P(D=\text{True}, C=\text{True}, A=\text{True}\ B=\text{False})$
$+ P(D=\text{True}, C=\text{True}, A=\text{False}, B=\text{False})$
$+ P(D=\text{True}, C=\text{True}, A=\text{False}, B=\text{True})$

$+ P(D=\text{True}, C=\text{False}\ A=T\ B=T)$
$P(D=\text{True}, C=\text{False}\ A=T\ B=F)$
$P(D=\text{True}, C=\text{False}\ A=F, B=F)$
$P(D=\text{True}, C=\text{False}\ A=F, B=T)$

$= (0.6)(0.4)(0.4)(0.7)$
$+ (0.6)\,(0.6)(0.4)(0.7)$
$+ (0.6)(0.7)(0.4)(0.7)$

$$P(A = F, C = T, E = T)$$

$$= \cancel{P(A) P(B) P(C}$$

$$= P(A = F)\Big[P(C = T, A = T, B = T) + P(C = T, A = T, B = F) + P(C = T, A = F, B = F) + P(C = T, A = T, B = F)\Big]$$
$$\Big[P(E = T, C = T) + P(E = T, C = F\Big]$$

— end of answersheet —