PRML Handwritten Assignment - 1

Mayank Dahiya
B23CS1035
Mayank Dahiya

Q-1 I affirm that this assignment is solely my work. I have not used unauthorised assistance, engaged in plagiarism, or voilated ethical standards. Further, all references used and any discussion with anyone have been appropriately cited. Any breach may lead to disciplinary actions as per the course academic honesty policy discussed in Lecture - 1.

Q-2 Positive Samples : $(1, +1)$, $(2, +1)$, $(5, +1)$ &
Negative samples : $(4, -1)$, $(6, -1)$, $(7, -1)$

Classification Region for test sample $(0 \leq x \leq 7)$

(a) (i) $[0, 3]$ → closest to 1 and 2 ⇒ classified as (+1) positive.

(ii) $[3, 4.5]$ → Closest to 4 ⇒ classified as ~~-ve~~ (-1) negative

(iii) $(4.5, 5.5]$ → Closest to 5 ⇒ classified as (+1) positive

(iv) $(5.5, 7]$ → closest to 6 and 7 ⇒ classified as (-1) negative

- $[0, 3]$ and $[4.5, 5.5]$ will be classified as +ve.

(b) Training accuracy for $k = 3$ in KNN.

Training point	3-Nearest Neighbours	Vote	Correct
1(+1)	2(+1), 4(-1), 5(+1)	(+1)	✓
2(+1)	1(+1), 4(-1), 5(+1)	(+1)	✓
4(-1)	2(+1), 5(+1), 6(-1)	(+1)	✗
5(+1)	4,(-1), 6(-1), 7(-1)	(-1)	✗
6(-1)	7(-1), 5(+1), 4(-1)	(-1)	✓
7(-1)	6(-1), 5(+1), 4(-1)	(-1)	✓

Total correctly classified points : 4/6

Training accuracy $= 4/6 \times 100 = 66.67\%$
(for $k = 3$)

Q-3 Query -1, it retrieves 10 images as follows -

-1, +1, +1, -1, +1, +1, -1, -1, +1, -1

+1 → retrieved image is relevant = 5

-1 → irrelevant to the query = 5

Total no. of relevant images in the database = 10.

(a) Precision $= \dfrac{\text{Relevant}}{\text{Total images retrieved}} = \dfrac{\text{True positive}}{\text{Total images retrieved}}$

$= \dfrac{5}{10} = 0.5$ ⇒ ~~Probability~~ Precision $= 50\%$

(b) Recall $= \dfrac{\text{True Positive (Relevant images retrieved)}}{\text{Total Relevant Images in database}}$

$= \dfrac{5}{10} = 0.5 \Rightarrow$ Recall $= 50\%$

Q-4 Total students - 100

RT - PCR tested positive : $\{S_1, S_2\}$

Actually positive students : $\{S_1, S_2, S_5, S_9, S_{10}, S_{11}, S_{100}\}$

(a) True Positive :→ (Student correctly tested positive)

TP → 2 $\{S_1, S_2\}$

(b) False Positive :→ Student incorrectly tested positive

⇒ FP → 0 = None.

(c) False Negative :→ Students actually positive but tested negative.

FN → 5 $\{S_5, S_9, S_{10}, S_{11}, S_{100}\}$

(d) True Negative :→ Students correctly tested negative

⇒ TN → 93 = Total negative - False negative

$= 98 - 5 = 93$

(e) Precision $= \dfrac{TP}{TP + FP}$

$= \dfrac{2}{2+0} = 1$

⇒ Precision $= 100\%$

↳ all students whose test classified as positive were indeed positive.

(f) Recall $= \dfrac{TP}{TP+FN} = \dfrac{2}{2+5} = 0.2857$

Recall $= 28.57\,\%$

Q-5 $\quad y = \bar{x}_1 x_2 x_3 + \bar{x}_1 \bar{x}_2 x_3 + x_1 x_2 \bar{x}_3 + x_1$

$\left(\begin{array}{c} T=0 \\ F=1 \end{array}\right)$

(a)

x_1	x_2	x_3	$\bar{x}_1 x_2 x_3$	$\bar{x}_1 \bar{x}_2 x_3$	$x_1 x_2 \bar{x}_3$	x_1	y
T	T	T	T(0)	T(0)	T(0)	T(0)	F(0)
T	T	F	T(0)	F(1)	T(0)	T(0)	F(1)
T	F	T	T(0)	T(0)	T(0)	T(0)	T(0)
T	F	F	F(1)	T(0)	T(0)	T(0)	F(1)
F	T	T	T(0)	T(0)	T(0)	F(1)	F(1)
F	T	F	T(0)	T(0)	T(0)	F(1)	F(1)
F	F	T	T(0)	T(0)	F(1)	F(1)	F(1)
F	F	F	T(0)	T(0)	T(0)	F(1)	F(1)

(b) <u>Decision Tree Construction</u>

Calculate the impurity measure for the root node and each of the candidate split criteria : $x_1 = 0$, $x_2 = 0$, $x_3 = 0$

Step-1 : Impurity calculation for root node

Total samples $= 8$.

$y = 1$ (6 samples)

$y = 0$ (2 samples)

Gini Impurity is :→

$$\text{Gini}(D) = 1 - \sum_{i=1}^{c} p_i^2$$

$$P_1(y=1) = \frac{6}{8} = 0.75$$

$$P_0(y=0) = \frac{2}{8} = 0.25$$

$$\text{Gini}(D) = 1 - (0.75^2 + 0.25^2)$$

$$= 1 - (0.5625 + 0.0625)$$

$$= 1 - 0.625$$

$$= 0.375$$

So, <u>Gini Impurity at root node</u> = 0.375

(ii) Misclassification impurity is :→

$$\text{Misclassification}(D) = 1 - \max(P_1, P_0)$$

$$\Rightarrow \text{Misclassification}(D) = 1 - \max(0.75, 0.25)$$

$$= 1 - 0.75 = 0.25$$

⇒ <u>Misclassification impurity at root node</u> = 0.25

(iii) Entropy impurity at root node

$$\text{Entropy}(D) = - \sum_{i=1}^{c} p_i \log_2(p_i)$$

$P_1 = 0.75$, $P_0 = 0.25$

$$\text{Entropy}(D) = -(0.75 \log_2 0.75 + 0.25 \log_2 0.25)$$

$$= -(-0.31128 - 0.5)$$

$$\boxed{\text{Entropy}(D) = 0.8113}$$

Gini Impurity $= 0.375$
Misclassification impurity $= 0.25$
Entropy Impurity $= 0.8113$

Split criteria :→ $x_1 = 0$, $x_2 = 0$, $x_3 = 0$

Split on $x_1 = 0$ vs $x_1 = 1$

$$y = \{0, 1, 0, 1\} \quad (2 \text{ zeros}, 2 \text{ ones})$$

Gini : $G_L = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$

Entropy : $H_L = -\left(\frac{2}{4}\log_2 \frac{2}{4} + \frac{2}{4}\log_2 \frac{2}{4}\right) = 1$

For Right $x_1 = 1$, $y = \{1, 1, 1, 1\}$

Gini Impurity $= 1 - 1 = 0$

Entropy : $H_R = 0$

Weighted Impurity :

$$G_{split} = \frac{n_1}{n} G_1 + \frac{n_2}{n} G_2$$

where –

n_1 = no. of samples in first child node after split

n_2 = no. of samples in second child node after split.

$n = n_1 + n_2$

G_1 = impurity for first child node

G_2 = impurity for second child node

$$\Rightarrow G_{split} = \frac{4}{8}(0.5) + \frac{4}{8}(0) = 0.25$$

Information gain $= IG = H_{root} - H_{split}$

$$H_{split} = \frac{n_1}{n} H_1 + \frac{n_2}{n} H_2$$

$H_1 \rightarrow$ Left entropy

$H_2 \rightarrow$ Right entropy

$$H_{split} = \frac{4}{8} \times 1 + \frac{4}{8} \times 0 \quad \cancel{= 0.25}$$
$$= 0.5$$

$\Rightarrow$ Information gain $(IG) = 0.8113 - 0.5$
$$= 0.3113$$

(ii) Repeat for $x_2 = 0$ and $x_2 = 1$

$x_2 = 0 \quad y = \{0, 1, 1, 1\}$

$x_2 = 1 \quad y = \{0, 1, 1, 1\}$

$$G_L = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2$$
$$= 1 - \frac{10}{16} = \frac{6}{16} = 0.375$$

$$G_R = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 = 0.375$$

Weighted Gini Impurity

$$G_{split} = \frac{4}{8} \times 0.375 + \frac{4}{8} \times 0.375$$

$$\Rightarrow G_{split} = 0.375$$

$$H_L = -\left(\frac{3}{4}\log_2\left(\frac{3}{4}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right)\right)$$
$$= 0.8113$$

$H_R = 0.8113$

$\Rightarrow H_{split} = 0.8113$ $\qquad \Rightarrow$ Information gain $= 0$

$\Rightarrow H_{split} = H_{Root}$

Split on x_3

$x_3 = 0 \quad y = \{0,0,1,1\}$

$x_3 = 1 \quad y = \{1,1,1,1\}$

$$G_L = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2$$

$$= 0.5$$

$$G_R = 1 - \left(\frac{4}{4}\right)^2 = 0$$

Weighted Gini impurity,

$$G_{split} = \frac{4}{8} \times 0.5 + \frac{4}{8} \times 0$$

$$= 0.25$$

$$H_L = -\left(\frac{2}{4}\log_2\left(\frac{2}{4}\right) + \frac{2}{4}\log_2\left(\frac{2}{4}\right)\right) = 1$$

$$H_R = 0$$

Weighted Entropy, $H_{split} = \frac{4}{8} \times 1 + \frac{4}{8} \times 0$

$$= 0.5$$

Information gain $= H_{root} - H_{split}$

$$= 0.8113 - 0.5$$

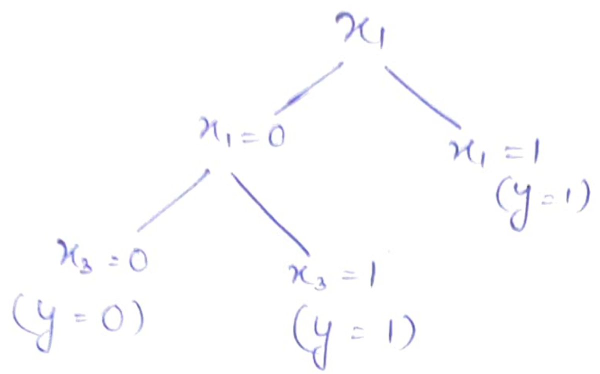
$$= \underline{0.3113}$$

Feature	Information gain
x_1	0.3113
x_2	0
x_3	0.3113

either can be choosen as root split

(iv) Decision Tree

Highest Information Gain = 0.3113 (x_1) as root node



(v) Classifying $x_1 = x_2 = x_3 = 0$

$x_1 = 0$, $x_2 = 0$, $x_3 = 0$

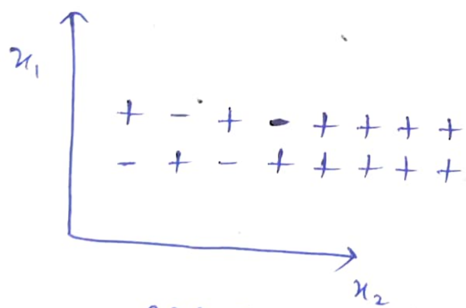
$x_1 = 0$ (move left)

$x_3 = 0$ (move left)

and $y = 0$

→ $x_3 = 0$, there are equal probability to choose $y = 0$ and $y = 1$ so $x_3 = 1$ $(y = 1)$ 100% probability so I choose $y = 0$ for $x_3 = 0$.

Q - 6



ABC Decision Tree

(a) Number of leaf nodes in the ABC decision tree :

→ Since the ABC algorithm doesnot prune, it will continue splitting until each region contains only one class.

If every unique combination of features leads to a pure node, then the number of leaf nodes corresponds to the number of unique patterns. (8 positive (+) and 8 negative (−))

→ The ABC ~~Leaf nodes~~ decision tree will have 16 leaf nodes.

(b) Leave - one out classification error for ABC

The leave - one out classification error involves training the decision tree on (n−1) samples and testing it on the left-out sample for an unpruned decision tree (ABC) → since ABC decision tree is fully grown, each training set (after removing one point) will still perfectly classify all training examples —

⇒ the total number of misclassification in leave - one - out cross - validation is 0.

(c) Leave - one out classification Error for XYZ Algorithm

The XYZ algorithm doesnot perform any splits, meaning it constructs a decision tree with a single leaf node. This means the entire dataset is classified under a single majority class.

Misclassification analysis →

→ Since there is only one leaf node, the algorithm will predict the majority class for all points.

→ 10 positive sample (+) and 6 negative sample (−)

→ Total number of misclassification = min (10, 6)
= 6.

(d) (i) ABC algorithm constructs an unpruned decision tree, splitting until each training sample is perfectly classified. This leads to overfitting because the tree memorizes the training data and may not generalize well to unseen data.

(ii) XYZ algorithm, on the other hand, avoids splitting entries resulting in underfitting.

⇒ ABC algorithm will overfitt to the training data.

Q-7 (a) Machine Learning :- A field of AI whose system learns pattern from data to make predictions or decisions without being explicit programs.

(b) Overfitting :- When a model learns the training data too well, including noise and performs poorly on unseen data.

(c) Supervised Learning :- A type of machine learning where models are trained on labeled data to predict outcomes for new inputs.

(d) CART Decision Tree :- A binary decision tree algorithm that uses Gini impurity or entropy to split data and produces classification or regression trees.

(e) ID3 Decision Tree :- A decision tree algorithm that uses information gain based on entropy to select the best feature for splitting at each node.

Q-8 Linear Regression Model $(-1,0), (0,-1), (2,1)$

(a) Assume $\hat{y}_i = b + n_i$, where n_i is gaussian noise

↳ mean of all y_i

$$\Rightarrow b = \frac{0 + (-1) + 1}{3} = 0$$

$\Rightarrow$ Estimated model is $\underline{\hat{y} = 0 + n_i}$

(b) Assume $\hat{y}_i = m x_i + n_i$

Best-fit slope m using least square regression

$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\bar{x} = \frac{-1+0+2}{3} = \frac{1}{3}, \quad \bar{y} = \frac{0-1+1}{3} = 0$$

$$m = \frac{\cancel{\left(-1-\frac{1}{3}\right)(0-0)} + \left(0-\frac{1}{3}\right)(-1-0) + \left(2-\frac{1}{3}\right)(1-0)}{\left(-1-\frac{1}{3}\right)^2 + \left(0-\frac{1}{3}\right)^2 + \left(2-\frac{1}{3}\right)^2}$$

$$= \frac{\frac{1}{3} + \frac{5}{3}}{\frac{16}{9} + \frac{1}{9} + \frac{25}{9}} = \frac{2}{\frac{42}{9}} = \frac{3}{7}$$

$\Rightarrow$ Estimated model is

$$\underline{\hat{y} = \frac{3}{7} x + n_i}$$

Q-9. $P(X/+) \sim U[3,6] \Rightarrow$ X given class(+) is uniformly distributed.

$P(X/-) \sim U[1,4] \Rightarrow$ X given class(−) is uniformly distributed.

$$P(+) = 0.8, \quad P(-) = 0.2$$

Computing likelihood functions

$$P(X/+) = \frac{1}{6-3} = \frac{1}{3}, \quad 3 \le X \le 6$$

$$P(X/-) = \frac{1}{4-1} = \frac{1}{3}, \quad 1 \le X \le 4$$

→ From Bayes Rule,

$$P(+/X) = \frac{P(X/-)\,P(+)}{P(X)}$$

$$P(-/X) = \frac{P(X/-)\,P(-)}{P(X)}$$

$$\Rightarrow P(X) = P(X/+)\,P(+) + P(X/-)\,P(-)$$

$P(+) = 0.8, \; P(-) = 0.2$ $\Big\{$ Overlap region : [3,4] non-zero probabilities in this region

→ Finding the decision boundary.

For X we compare,

$$P(+/X) > P(-/X) \Rightarrow P(X/+)\,P(+) > P(X/-)\,P(-)$$

$$\Rightarrow \frac{1}{\cancel{3}} \times 0.8 > \frac{1}{\cancel{3}} \times 0.2$$

$$\Rightarrow 0.8 > 0.2$$

→ We assign X to class + whenever both decision boundary occurs where the likelihoods change, which is at $x = 3$ and $x = 4$.

→ Error Calculation

Error happens when we incorrectly classify points

1) Region $[1,3)$: Misclassify class (−) and predict (+), we always predict (+).

$$\text{Error} = P(x \in [1,3) \mid -)\, P(-) = \frac{\text{Length of region}}{\text{Total length of } P(x \mid -)}$$

Interval = 2

$$= \frac{2}{3}$$

$$P(\text{misclassify}) = \frac{2}{3} \times 0.2 = \frac{2}{15}$$

2) Region $[3,4]$: Misclassify (−) as class (+)

$$\text{Error} = \frac{\text{Length of region}}{\text{Total length}} \times P(-)$$

$$= \frac{1}{3} \times 0.2 = \frac{1}{15}$$

3) Region $(4,6]$: No error

Only class (+) exist and we correctly predict it.

→ Total probability error $= \frac{2}{15} + \frac{1}{15} = \frac{1}{5} = 0.2 = 20\%$

Q-10 (a) The covarience matrix is a square matrix that represent the covarience between multiple random variables.

For a random vector X with n variables:

$$\Sigma = E[(X - E[X])(X - E[X])^T]$$

where $\Sigma_{ij} = Cov(X_i, X_j)$

→ Properties of the Covariance matrix.

1) Symmetry →

$$\Sigma_{ij} = \Sigma_{ji}$$

the covariance matrix is always symmetric.

2) Positive - Semi - Definiteness →

$$v^T \Sigma v \geq 0 \quad \forall \; v \in R^n$$

All ~~eigenvectors~~ eigenvalues of Σ are non-negative.

3) Diagonal elements represent variances ÷

$$\Sigma_{ii} = Var(X_i)$$

The diagonals elements of Σ are the variance of individual variables.

Q-4(b) Given: (1,2), (3,4), (4,5), (5,5), (5,6), (10,10)

To find mean vector u

The ~~g~~ mean vector is given by $u = \frac{1}{n}\sum_{i=1}^{n} x_i$

where x_i are the data points, $n = 6$.

→ Mean of x_1 values

$$u_1 = \frac{1+3+4+5+5+10}{6} = 4.67$$

→ Mean of x_2 values

$$u_2 = \frac{2+4+5+5+6+10}{6} = 5.33$$

Thus, mean vector is:

$$u = \begin{bmatrix} 4.67 \\ 5.33 \end{bmatrix}$$

→ For the covariance matrix,

$$\Sigma = \frac{1}{n} \sum_{i=1}^{n} (x_i - u)(x_i - u)^T$$

where $(x_i - u)$ is the deviation from mean.

X_1	X_2	$X_1 - u_1$	$X_2 - u_2$
1	2	-3.67	-3.33
3	4	-1.67	-1.33
4	5	-0.67	-0.33
5	5	0.33	-0.33
5	6	0.33	0.67
10	10	5.33	4.67

The covariance formula is:

$$\mathrm{cov}(x_i, x_j) = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - u_i)(x_{kj} - u_j)$$

$$a_{11} = \frac{1}{5}\left((-3.67)^2 + (-1.67)^2 + (-0.67)^2 + (0.33)^2 + (0.33)^2 + (5.33)^2 \right)$$

$$\Rightarrow \underline{a_{11} = 9.07}$$

$$a_{22} = \frac{1}{5}\left((-3.33)^2 + (-1.33)^2 + (-0.33)^2 + (-0.33)^2 + (0.67)^2 + (4.67)^2 \right)$$

$$\Rightarrow \underline{a_{22} = 7.07}$$

$$a_{12} = a_{21} = \frac{1}{5}\left[(-3.67)(-3.33) + (-1.67)(-1.33) + (-0.67)(-0.33) + (0.33)(-0.33) + (0.33)(0.67) + (5.33)(4.67)\right]$$

$$\Rightarrow \underline{a_{12} = a_{21} = 7.93}$$

$$\rightarrow \Sigma = \begin{bmatrix} 9.07 & 7.93 \\ 7.93 & 7.07 \end{bmatrix}$$

Q-12 Datasets of Fruit (either apple or oranges) with two features.

Colour (Red or Orange)

Shape (Round or not round)

The following probabilities are given :

$P(\text{apple}) = 0.6$ $\qquad$ $P(\text{RED} \mid \text{apple}) = 0.8$

$P(\text{orange}) = 0.4$ $\qquad$ $P(\text{Round} \mid \text{apple}) = 0.7$

$P(\text{RED} \mid \text{orange}) = 0.6$ $\qquad$ $P(\text{Round} \mid \text{orange}) = 0.3$

→ For posterior probabilities for both classes (apple and orange)

From Bayes Theorum :

$$P(\text{Class} \mid \text{Data}) = \frac{P(\text{Data} \mid \text{Class}) \times P(\text{Class})}{P(\text{data})}$$

We are calculating : $P(\text{apple} \mid \text{Red, Round})$

$P(\text{orange} \mid \text{Red, Round})$

→ Calculate the numerator for both classes -

For apple →

$$P(\text{apple} \mid \text{Red, Round}) \propto P(\text{Red} \mid \text{Apple}) \times P(\text{Round} \mid \text{Apple}) \times P(\text{apple})$$

$$P(\text{apple} \mid \text{Red, Round}) \propto 0.8 \times 0.7 \times 0.6$$

$$\propto 0.336$$

(Denominator can be ignored as it is same)

For orange →

$P(\text{Orange} | \text{Red}, \text{Round}) \propto P(\text{Red} | \text{orange}) \times P(\text{Round} | \text{Orange}) \times P(\text{orange})$

$\propto 0.6 \times 0.3 \times 0.4 = 0.072$

Since $P(\text{apple} | \text{Red}, \text{Round}) > P(\text{orange} | \text{Red}, \text{Round})$

We can conclude that fruit is more likely to be an apple.

Q-13 Given discrete random variable X with support $\{-2, -1, 0, 1, 2\}$ it follows uniform distribution

To find the expected value and varience for X.

Uniform Distribution $\{-2, -1, 0, 1, 2\}$

PMF ⇒ $P(X=x) = \frac{1}{5}$ for $x \in \{-2, -1, 0, 1, 2\}$.

Expected value $E[X] = \sum_{x \in \{-2,-1,0,1,2\}} x \cdot P(X=x)$

$\Rightarrow E[X] = (-2) \times \frac{1}{5} + (-1) \frac{1}{5} + 0 \times \frac{1}{5} + 1 \times \frac{1}{5} + 2 \times \frac{1}{5}$

$\Rightarrow \underline{E[X] = 0}$

$Var(X) = E[X^2]$

$E[X^2] = \sum_{x \in (-2,-1,0,1,2)} x^2 \cdot P(X=x)$

$\Rightarrow E[X^2] = (-2)^2 \times \frac{1}{5} + (-1)^2 \times \frac{1}{5} + 0^2 \times \frac{1}{5} + 1^2 \times \frac{1}{5} + 2^2 \times \frac{1}{5}$

$\Rightarrow \underline{E[X^2] = 2}$

$\Rightarrow Var(X) = 2.$

Q-14. Given: Category 1:

$$P(x|1) \sim N(0,1)$$

$\hookrightarrow$ mean = 0, covarience is identity

Category 2:

$$P(x|2) \sim N\left(\begin{bmatrix}1\\1\end{bmatrix}, I\right)$$

$\hookrightarrow$ mean = (1,1), covarience = I

Category 3:

$$P(x|3) = \frac{1}{2} N\left(\begin{bmatrix}0.5\\0.5\end{bmatrix}, I\right) + \frac{1}{2} N\left(\begin{bmatrix}-0.5\\-0.5\end{bmatrix}, I\right)$$

$\hookrightarrow$ mixture of two normal distributions

Each class has an equal prior probability

$$P(i) = \frac{1}{3}, \quad i = 1,2,3$$

(a) Classification of $x = (0.3, 0.3)$ using minimum error criterion

Using gaussian density function:

$$P(x|i) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x-u_i)^T I^{-1}(x-u_i)\right)$$

→ Computing likelihoods:

For class 1 $(u_1 = (0,0))$

$$P(x|1) = \frac{1}{2\pi} e^{-\frac{1}{2}((x-u_i)^T I^{-1}(x-u_i))}$$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}(0.3^2+(0.3)^2)} = \underline{\frac{1}{2\pi} e^{-0.09}}$$

(b) For class 2: $(u_2 = (1,1))$

$$P(x|2) = \frac{1}{2\pi} e^{\left(-\frac{1}{2}((0.3-1)^2 + (0.3-1)^2)\right)}$$

$$= \underline{\frac{1}{2\pi} e^{-0.49}}$$

(c) For class 3 ($u_3 = (0.5, 0.5)$ and $(-0.5, -0.5)$)

$$P(x|3) = \frac{1}{2}\left[\frac{1}{2\pi} e^{-0.08} + \frac{1}{2}\left[\frac{1}{2\pi} e^{-0.64}\right]\right]$$

$$P(x|1) > P(x|3) > P(x|2)$$

Since Class 1 has the highest posterior probability, for point $x = [0.3, 0.3]$

2] Classification when first feature is missing $x = (*, 0.3)$

$$P(x_2|i) = \int P(x_1, x_2|i)\, dx_1$$

The point $x = [*, 0.3]$ is classified as belonging to category 1 as it has the highest posterior probability.

3] Classification when second feature is missing $x = (0.3, *)$

$P(1|x_1 = 0.3) \approx 0.3688$ → highest posterior

$P(2|x_1 = 0.3) \approx 0.3026$ ⇒ missing belong to class 1.

$P(3|x_1 = 0.3) \approx 0.3292$

Q-15.

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

Total number of sample = 8

For $k = 1 \longrightarrow 4$ time

$k = 0 \longrightarrow 4$ time

1) $P(K=1) = \dfrac{\text{no. of way of } k=1}{\text{Total}}$

$= \dfrac{4}{8} = 0.5$

$P(K=0) = \dfrac{4}{8} = 0.5$

2) For $P(a=1|K)$, $P(b=1|K)$, $P(c=0|K)$

1) Given $K=1$:

$a=1$ appears 2 times (out of 4)

$\Rightarrow P(a=1|k=1) = \dfrac{2}{4} = 0.5$

$b=1$ appears 1 time

$\Rightarrow P(b=1|k=1) = \dfrac{1}{4} = 0.25$

$c=0$ appears 0 times

$\Rightarrow P(c=0|k=1) = 0.$

2) when $k=0$:

$a=1$ appears 2 time (out of 4)

$\Rightarrow P(a=1|k=0) = \dfrac{2}{4} = 0.5$

$b=1$ appears 2 times

$\Rightarrow P(b=1|k=0) = \dfrac{2}{4} = 0.5$

$c=0$ appears 1 time

$\Rightarrow P(c=0|k=0) = \dfrac{1}{4} = 0.25$

(a) Compute $P(K=1 \mid a=1, b=1, c=0)$ using Naive Bayes Classifier

$$P(K=1 \mid a=1, b=1, c=0) \propto P(a=1 \mid k=1)\, P(b=1 \mid k=1)\, \cancel{P(c=0 \mid k=1)}^{\,0}\, P(k=1)$$

$$\Rightarrow P(k=1 \mid a=1, b=1, c=0) = 0$$

(b) Compute $P(k=0 \mid a=1, b=1)$

$$P(a=1, b=1 \mid k=0) = P(a=1 \mid k=0) \times P(b=1 \mid k=0)$$

$$= 0.5 \times 0.5$$

$$= 0.25$$

$$P(k=0 \mid a=1, b=1) = \frac{P(a=1, b=1 \mid k=0)\, P(k=0)}{P(a=1, b=1)}$$

$$P(k=0) = 0.5$$

$$P(a=1, b=1, k=0) = 0.25$$

$$P(a=1, b=1 \mid k=1) = P(a=1 \mid k=1) \times P(b=1 \mid k=1)$$

$$= 0.5 \times 0.25$$

$$= 0.125$$

To find $P(a=1, b=1) = P(a=1, b=1 \mid k=0)\, P(k=0) + P(a=1, b=1 \mid k=1)\, P(k=1)$

$$= 0.25 \times 0.5 + 0.125 \times 0.5$$

$$= 0.125 + 0.0625$$

$$= 0.1875$$

$$\Rightarrow P(K=0 \mid a=1, b=1) = \frac{0.25 \times 0.5}{0.1875} = \frac{2}{3}$$

(c) $P(k=1 \mid a=1, b=1, c=0)$

$$P(k=1 \mid a=1, b=1, c=0) = \frac{\cancelto{0}{P(a=1, b=1, c=0 \mid k=1)}}{P(a=1, b=1, c=0)}$$

$$= 0.$$

(d) For $P(k=0 \mid a=1, b=1)$,

$$P(k=0 \mid a=1, b=1) = \frac{P(a=1, b=1 \mid k=0)}{P(a=1, b=1)}$$

$(a=1, b=1)$ appears twice, once with $k=0$ and once with $k=1$

$$\Rightarrow B = \frac{1}{2}$$

$$\Rightarrow P(k=0 \mid a=1, b=1) = 0.5$$

<u>Answer</u> :

(a) $P(k=1 \mid a=1, b=1, c=0) = 0$

(b) $P(k=0 \mid a=1, b=1) = 2/3$

(c) $P(k=1 \mid a=1, b=1, c=0) = 0$

(d) $P(k=0 \mid a=1, b=1) = \frac{1}{2}$