Assignment 2 ,Machine Learning,CS-GY 6923, Th 6:00 - 8:30 PM

Shantanu Tripathi

[Note: Do not consider the page numbers written on the top right corner of the page. They were just for the ease of arranging images.]

1

Homework 2

Ans 1) a)

$$p(\text{pos. test} \mid \neg d) = 0.2$$

$$p(\text{-ve test} \mid d) = 0.15$$

$$p(\text{pos.test} \mid d) = 1 - p(\text{-ve test} \mid d)$$
$$= 1 - 0.15$$
$$= 0.85$$

$$p(\text{-ve test} \mid \neg d) = 1 - p(\text{pos. test} \mid \neg d)$$
$$= 1 - 0.2$$
$$= 0.8$$

$$p(d) = 0.01\,\%$$
$$= \frac{0.01}{100} = 0.0001$$

$$p(\neg d) = 1 - p(d) = 0.9999$$

$$p(\text{disease} \mid \text{+ve test}) = \frac{p(\text{pos.test} \mid d) \cdot p(d)}{p(\text{+ve test})}$$

1

$\Rightarrow$ ~~p(+ve test)~~

~~$= p(\text{pos. test} \mid d) \cdot p(d) + p(\text{pos test} \mid \neg d) \cdot p(\neg d)$~~

$\Rightarrow p(\text{+ve test})$

$= p(\text{pos. test} \mid d) \cdot p(d) + p(\text{pos. test} \mid \neg d) \cdot p(\neg d)$

$= 0.85 \times 0.0001 + 0.2 \times 0.9999$

$= 0.200065$

$\Rightarrow \boxed{p(\text{+ve test}) = 0.200065}$

Putting values in eq. 1

$$p(d \mid \text{+ve test}) = \frac{0.85 \times 0.0001}{0.200065}$$

$= 4.25 \times 10^{-4}$

$= 0.000425$

$\therefore \boxed{p(d \mid \text{+ve test}) = 0.000425}$

~~$\therefore p($~~

Here,

$\therefore p(\neg d \mid +ve\ test) = 0.99957$

$\therefore \boxed{p(\neg d \mid +ves\ test) > p(d \mid +ve\ test)}$

We predict that the person $\boxed{\text{doesn't have the disease}}$

Ans 1) b)

In ML Hypothesis, we do not have a prior.

$\cancel{Y_{ML} = \arg\max_{y} p(\text{pos}}$

$$Y_{ML} = \arg\max_{y} p(\text{pos test} \mid y)$$

$$= \max\left(p(\text{pos. test} \mid d),\ p(\text{pos. test} \mid \neg d)\right)$$

$$= \max(0.85, 0.2)$$

$\therefore p(\text{pos test} \mid d) > p(\text{pos. test} \mid \neg$

we have the conculesion that the person $\boxed{\text{has the disease.}}$

Ans 1) c)

$p(\text{+ve test} \mid \neg d) = 0.07$

$p(\text{-ve test} \mid d) = 0.04$

$p(\text{+ve test} \mid d) = 0.96$

$p(\text{-ve test} \mid \neg d) = 0.93$

$$\Rightarrow \boxed{p(d \mid \text{+ve test}) = \frac{p(\text{+ve test} \mid d) \cdot p(d)}{p(\text{+ve test})}}$$

$\Rightarrow$ $p(\text{+ve test}) =$

$p(\text{+ve test} \mid d) \cdot p(d) +$

$p(\text{+ve test} \mid \neg d) \cdot p(\neg d)$

$= 0.96 \times 0.0001 + 0.07 \times 0.9999$

$= 0.070089$

$\therefore p(d \mid +ve\ test)$

$$= \frac{0.96 \times 0.0001}{0.070089}$$

$$= 0.00137$$

$\therefore \boxed{p(d \mid +ve\ test) = 0.00137}$

To ~~cla~~ calculate the complete posterior probability for the patient to have the disease:

$P(disease) = P(test_1 \cap test_2)$

$= P(test_1) \cdot P(test_2)$ $\because$ tests are independen

$= 0.000425 \times 0.00137$

$= 5.82 \times 10^{-7}$

~~Total prob. = 5.82 × 10^-8~~

$$\boxed{\text{Total prob.} = 5.82 \times 10^{-7}}$$

Ans 2)

a) for the values of x_2,

$X_{min} = -1$

$X_{max} = 51$

formula for scaling =>

$$\Rightarrow \frac{X - X_{min}}{X_{max} - X_{min}}$$

$$= \frac{X - (-1)}{51 - (-1)} = \cancel{\frac{X+1}{52}} \quad \frac{X+1}{52}$$

$\Rightarrow$

X_2	Scaled Value
40	41/52 = 0.79
51	52/52 = 1.00
-1	0/52 = 0.00
2	3/52 = 0.06
26	27/52 = 0.52
41	42/52 = 0.81

Ans 2) b)

$$x = \begin{bmatrix} 3.9 \\ 4 \end{bmatrix}$$

$$x_scaled = \begin{bmatrix} \frac{3.9 - (-0.3)}{3.8 - (-0.3)} \\ \frac{4+1}{52} \end{bmatrix}$$

$$\cancel{= \begin{bmatrix} \frac{4.2}{4.1} \\ 5 \end{bmatrix}}$$

$$= \begin{bmatrix} 4.2/4.1 \\ 5/52 \end{bmatrix}$$

$$x_scaled = \begin{bmatrix} 1.02 \\ 0.09 \end{bmatrix}$$

$x_scaled = \begin{bmatrix} 1.02 \\ 0.09 \end{bmatrix} \begin{matrix} \rightarrow x_a \\ \rightarrow x_b \end{matrix}$

X_1	X_2	$a = X_1 - x_a$ (1.02)	$b = X_2 - x_b$ (0.09)	$a^2 + b^2$
0.68	0.79	−0.34	0.70	0.6056
1.00	1.00	−0.02	0.91	0.8285
0.00	0.00	−1.02	−0.09	1.0485
0.24	0.06	−0.78	−0.03	0.6093
0.46	0.52	−0.56	0.43	0.4985
0.63	0.81	−0.39	0.72	0.6705

Equcledian distance wants the square root of $a^2 + b^2$

eucled dist
0.77820
0.9102
1.023
0.7805
0.7060
0.8188

minimum value

$\boxed{\text{Class : } -ve}$ as its closest to a point of −ve class

Ans 3)

Multivariate Guassian parameter derivation for "+" label

X1	X2
2.7	4.8
3.2	5.1
-0.4	-0.3
5.5	9.6

$$\mu = \begin{bmatrix} 5.5/3 \\ 9.6/3 \end{bmatrix} = \begin{bmatrix} 1.833 \\ 3.2 \end{bmatrix}$$

$$\text{Let } X = \begin{bmatrix} 2.7 & 4.8 \\ 3.2 & 5.1 \\ -0.4 & -0.3 \end{bmatrix}$$

$$X - \mu = \begin{bmatrix} 2.7 - 1.833 & 4.8 - 3.2 \\ 3.2 - 1.833 & 5.1 - 3.2 \\ -0.4 - 1.833 & -0.3 - 3.2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.867 & 1.6 \\ 1.367 & 1.9 \\ -2.233 & -3.5 \end{bmatrix}$$

$$\Sigma = (x-\mu)^T (x-\mu) \times (1/N)$$

$$= \begin{bmatrix} 0.867 & 1.367 & -2.233 \\ 1.6 & 1.9 & -3.5 \end{bmatrix} \begin{bmatrix} 0.867 & 1.6 \\ 1.6367 & 1.9 \\ -2.233 & -3.5 \end{bmatrix}$$

$$= \frac{\begin{bmatrix} (0.867)^2 + (1.367)^2 + (2.233)^2 & 0.867 \times 1.6 + \dots \\ 0.867 \times 1.6 + 1.36 \times 1.9 + 3.5 \times 2.23 & (1.6)^2 + (1.9)^2 + (3.5)^2 \end{bmatrix}}{N}$$

$$\Sigma = \begin{bmatrix} 7.6067/3 & 11.8/3 \\ 11.8/3 & 18.42/3 \end{bmatrix}$$

Mult. Derivation for "6" - "9" label

X1	X2	$x1 - \mu_0$	$x2 - \mu_1$
0.6	0.5	-0.9	-2.033
1.8	2.8	0.3	0.267
2.1	4.3	0.6	1.767
4.5	7.6		

$$\mu = \begin{bmatrix} 1.5 \\ 2.533 \end{bmatrix}$$

$$\cancel{\Sigma_{11} = [(0.9)^2 + (0.3)^2 + (0.6)^2]/3}$$

$$\Sigma_{11} = \frac{\left((0.9)^2 + (0.3)^2 + (0.6)^2\right)}{3}$$

$$= \frac{1.26}{3} = 0.42$$

$$\Sigma_{12} = \frac{0.9 \times 2.033 + 0.3 \times 0.267 + 0.6 \times 1.767}{3}$$

$$= \frac{2.97}{3} = 0.99$$

$$\Sigma_{21} = \Sigma_{12}$$

$$\Sigma_{22} = \frac{(2.033)^2 + (0.267)^2 + (1.767)^2}{3}$$

$$= \frac{7.326}{3} = 2.4422$$

$$\Rightarrow \Sigma = \begin{bmatrix} 0.42 & 0.99 \\ 0.99 & 2.442 \end{bmatrix}$$

$$\Rightarrow \mu_+ = \begin{bmatrix} 1.833 \\ 3.2 \end{bmatrix}$$

$$\Sigma_+ = \begin{bmatrix} 2.535 & 3.933 \\ 3.933 & 6.14 \end{bmatrix}$$

$$\Rightarrow \mu_- = \begin{bmatrix} 1.5 \\ 2.533 \end{bmatrix}$$

$$\Sigma_- = \begin{bmatrix} 0.42 & 0.99 \\ 0.99 & 2.4422 \end{bmatrix}$$

$\Rightarrow$ General formula for gaussian :

$$\frac{1}{\sqrt{2\pi|\Sigma|}} \cdot \exp\left(-\frac{1}{2}(x-\mu)^T |\Sigma|^{-1} \cdot (x-\mu)\right)$$

Note, here X is a column vector i.e.

$$\cancel{X = \begin{bmatrix} \text{val of attribute 1} \\ \text{val of attribute 2} \end{bmatrix}}$$

$$X = \begin{bmatrix} \text{val. of attr. 1} \\ \text{val. of attr. 2} \\ \vdots \\ \text{val of attr. n} \end{bmatrix}$$

But, while calculating ~~μ_+ & Σ_{+}~~ μ_+ & Σ_+, X was considered as a row vector, i.e.

$X = [\text{val of attribut 1} \quad \text{val. of att. 2} \quad \ldots \text{val of } n]$

$$\Rightarrow \boxed{p(X \mid \text{label} = +ve) = \underbrace{p(X; \mu_+, \Sigma_+)}_{\text{value given by p.d.f. of +ve label.}}}$$

$$\Rightarrow \cancel{p(X \mid \text{label} = -ve) = p(X; \mu_{-ve}, \Sigma_-)}$$

$$\Rightarrow \boxed{p(X \mid \text{label} = -ve) = p(X; \underbrace{\mu_{-ve}, \Sigma_{-ve}}_{\text{Value given by p.d.f of -ve label}})}$$

$$p(\text{label} = +ve) = \frac{3}{6} = \frac{1}{2} = 0.5$$

$$p(\text{label} = -ve) = \frac{3}{6} = \frac{1}{2} = 0.5$$

Using MAP hypothesis

$$\cancel{y_{MAP} = \max_{\theta} (}$$

If,

$$\boxed{\frac{p(x \mid \text{label} = +ve) * p(+ve)}{p(x)} > \frac{p(x \mid \text{label} = -ve) \cdot p(-ve)}{p(x)}}$$

↳ then label is +ve,
otherwise label is -ve.

Ans 3) b)

$$X = \begin{bmatrix} 3.1 \\ 1.67 \end{bmatrix}$$

$$p(X \mid \text{label} = +ve)$$

$$= p(X; \mu_+, \Sigma_+)$$

$$= \frac{1}{\sqrt{2\pi |\Sigma_+|}} \cdot \exp\left(-\frac{1}{2} \cdot (X - \mu_+)^T \Sigma_+^{-1} \cdot (X - \mu_+)\right)$$

$$\Rightarrow \quad \Sigma_+ = \begin{bmatrix} 2.535 & 3.933 \\ 3.933 & 6.14 \end{bmatrix}$$

$$|\Sigma_+| = 0.096411$$

$$\Rightarrow \quad \mu_+ = \begin{bmatrix} 1.833 \\ 3.2 \end{bmatrix}$$

$$\Rightarrow \quad X - \mu_+ = \begin{bmatrix} 1.267 \\ -1.5 \end{bmatrix}$$

$$\Rightarrow \Sigma_+^{-1} = \frac{1}{|\Sigma_+|}\begin{bmatrix} 6.14 & -3.933 \\ -3.933 & 2.535 \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \frac{1}{0.09641}\begin{bmatrix} 6.14 & -3.933 \\ -3.933 & 2.535 \end{bmatrix}$$

$$\Sigma_+^{-1} = \begin{bmatrix} 51.17 & -32.775 \\ -32.775 & 21.125 \end{bmatrix}$$

$$\Rightarrow (x-\mu_+)^T \cdot \Sigma_+^{-1} \cdot (x-\mu_+)$$

$$= \begin{bmatrix} 1.267 & -1.5 \end{bmatrix}\begin{bmatrix} 51.17 & -32.775 \\ -32.775 & 21.125 \end{bmatrix}\begin{bmatrix} 1.267 \\ -1.5 \end{bmatrix}$$

$$= \cancel{\begin{bmatrix} 113.99 & -73.213 \end{bmatrix}}$$

$$= \begin{bmatrix} 113.99 & -73.213 \end{bmatrix}\begin{bmatrix} 1.267 \\ -1.5 \end{bmatrix}$$

$$= \begin{bmatrix} 254.244 \end{bmatrix}$$

$$\Rightarrow \exp\left(-\frac{1}{2} \times 254.244\right)$$

$$= \exp(-127.12)$$

$$\text{const. term} = \frac{1}{\sqrt{2\pi(0.12)}}$$

$$= 0.892$$

$$\Rightarrow \boxed{p(x \mid \text{label} = +ve) = 0.892 \cdot \exp(-127.12)}$$

$$\Sigma_+^{-1} = \frac{1}{0.09641}\begin{bmatrix} 6.14 & -3.933 \\ -3.933 & 2.535 \end{bmatrix}$$

$$\Sigma_+^{-1} = \begin{bmatrix} 63.685 & -40.794 \\ -40.794 & 26.294 \end{bmatrix}$$

$$\Rightarrow (x - \mu_+)^T \cdot \Sigma_+^{-1} \cdot (x - \mu_+)$$

$$= \begin{bmatrix} 1.267 & -1.5 \end{bmatrix} \begin{bmatrix} 63.685 & -40.794 \\ -40.794 & 26.294 \end{bmatrix} \begin{bmatrix} 1.267 \\ -1.5 \end{bmatrix}$$

$$= \begin{bmatrix} 141.879 & -101.657 \end{bmatrix} \begin{bmatrix} 1.267 \\ 1.5 \end{bmatrix}$$

$$= \begin{bmatrix} 332.23 & -91.126 \end{bmatrix} \begin{bmatrix} 1.267 \\ -1.5 \end{bmatrix}$$

$$\begin{bmatrix} 141.879 & -91.126 \end{bmatrix} \begin{bmatrix} 1.267 \\ -1.5 \end{bmatrix}$$

$$= 316.45$$

$$\Rightarrow \exp\left(-\frac{1}{2} \cdot (316.45)\right)$$

$$= \exp(-158.23)$$

~~$\Rightarrow$ const. te~~

$$\Rightarrow \text{const. term} = \frac{1}{\sqrt{2\pi(0.09641)}}$$

$$= 1.285$$

$$\Rightarrow \cancel{p(x \mid \text{label} = +ve) = 1.285 \cdot \exp(-127}$$

$$\Rightarrow \boxed{p(x \mid \text{label} = +ve) = 1.285 \times \exp(-158.23)}$$

Calculating $p(x \mid \text{label} = -ve)$

$$\mu_- = \begin{bmatrix} 1.5 \\ 2.533 \end{bmatrix}$$

$$\Sigma_- = \begin{bmatrix} 0.42 & 0.99 \\ 0.99 & 2.4422 \end{bmatrix}$$

$$|\Sigma_-| = 0.0486$$

$$\Sigma_-^{-1} = \frac{1}{|\Sigma_-|} \begin{bmatrix} 2.4422 & -0.99 \\ -0.99 & 0.42 \end{bmatrix}$$

$$\Sigma_-^{-1} = \begin{bmatrix} 53.52 & -21.69 \\ -21.69 & 9.205 \end{bmatrix}$$

$$x - \mu = \begin{bmatrix} 1.6 \\ -0.833 \end{bmatrix}$$

$$\Rightarrow (x-\mu)^T \cdot \Sigma_-^{-1} \cdot (x-\mu)$$

$$= \begin{bmatrix} 1.6 & -0.833 \end{bmatrix} \begin{bmatrix} 53.52 & -21.69 \\ -21.69 & 9.205 \end{bmatrix} \begin{bmatrix} 1.6 \\ -0.833 \end{bmatrix}$$

$$= \begin{bmatrix} 103.69 & -42.37 \end{bmatrix} \begin{bmatrix} 1.6 \\ -0.833 \end{bmatrix}$$

$$= 201.26$$

$$\Rightarrow \exp\left(-\frac{1}{2} \cdot 201.26\right)$$

$$= \exp(-100.63)$$

$\Rightarrow$ Const. term $= 1.86$

$$\Rightarrow \boxed{p(x \mid \text{label} = -ve) = 1.86 \exp(-100.63)}$$

So, the values we have:

$$p(x \mid \text{label} = +ve) = 1.285 \exp(-158.23)$$

$$p(x \mid \text{label} = -ve) = 1.86 \exp(-100.63)$$

$$\boxed{\begin{aligned} p(x) = {} & p(x \mid \text{label} = ve) \cdot p(+ve) \\ & + p(x \mid \text{label} = -ve) \cdot p(-ve) \end{aligned}}$$

$$p(+ve) = \frac{3}{6} = p(-ve)$$

$$\therefore \boxed{p(+ve) = p(-ve) = 0.5}$$

~~Bayesian~~

According to MAP hypothesis, label predicted for $x = +ve$ if:

$$\frac{p(x \mid label = +ve) \cdot p(+ve)}{p(x)} > \frac{p(x \mid label = -ve) \cdot p(-ve)}{p(x)}$$

(greater)

$\because$ $p(+ve) = p(-ve)$

& $p(X)$ is same for both

we are just left with:

~~$p(x \mid label = +ve) > p(x \mid label = -ve)$~~

~~$p(x \mid label) \overset{?}{\underset{0}{}} p$~~

$p(x \mid label = +ve) \overset{?}{\underset{0}{}} p(x \mid label = -ve)$

$1.285 \cdot e^{-158.23} \overset{?}{\underset{0}{}} 1.86 \cdot e^{-100.63}$

i.e.

$1.285 \cdot e^{-158.23} < 1.86\, e^{-100.63}$

$\because$ $p(X \mid label = -ve)$ dominates

i.e. $\boxed{\text{Class of } X \text{ is } -ve}$

Ans 3) c)

Advantage :)

We just have to calculate one covariance matrix instead of k.

~~but~~

Earlier, for 1 class we had to calculate: ~~d. (+1)~~

$$\frac{d \times (d+1)}{2} \text{ params.}$$

For k classes:

$$\frac{k \times d \times (d+1)}{2}$$

$\Rightarrow$ Now, since we have 1 cov. matrix, we have to just calculate:

$$\frac{d \times (d+1)}{2} \text{ params}$$

Advantage: Since no. of params to train are reduced, the time to train model reduces.

$k \cdot d \cdot \frac{(d+1)}{2}$ reduces to $\frac{d \cdot (d+1)}{2}$

where k: No. of classes
d: No. of dimensions

Disadvantage:

Covariance matrix gives a sense of relationship b/w different attributes.

For eg.

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

gives a relationship b/w 1st & 2nd attribute.

For a class, this relationship can vary.

For eg. Imagine we had a problem where we had to classify a shopkeeper as good or bad. The attributes & ~~of~~ dataset is as follows:

Quantity of items purchased	Avg. price per item	Shopkeeper type
3	50	Good
6	75	Good
4	100	Bad
8	300	Bad

=> Here, its observed that the good shopkeeper tends to give discounts if the customer ~~buys~~ buys a lot of products.

~~i.e => A good customer has~~

~~an~~

Disadvantage: Cov. matrix gives a sense of relationship between diff. Attributes. This might be different for different classes. This data is lost if we make a single cov. matrix.

~~for eg.~~

i.e. for a good shopkeeper,
the relationship b/w
Quantity & Avg. price
is inversely prop.

For a bad ~~say~~ shopkeeper,
the more products you
buy, bad shopkeeper
charges you more :

Quantity $\propto$ Avg price

In different classes,
Quantity & Avg. Price are
~~different~~ related in a different
way to each other ∴
its better to have 2
different Co-Variance Matrice

Ans 4)

a)

x1	x 2	x 3
L	N	R
M	N	G
L	N	G

} +ve data

$m = 0.2$

$\Rightarrow \cancel{P(x_1 = L \mid +) = N}$

Smoothing formula:

$\Rightarrow \cancel{N \times m}$ $\dfrac{t + m}{N + m \times S}$

$\Rightarrow P(x_1 = L \mid +) \Rightarrow$

$t = 2$, $m = 0.2$

$N = 3$ $S = 3$ $\{L, M, H\}$

$\Rightarrow \dfrac{2 + 0.2}{3 + 0.2 \times 3} = \dfrac{2.2}{3.6}$

$= 0.611$

$$\Rightarrow \boxed{P(x_1 = L \mid +) = 0.611}$$

$$P(x_2 = Y \mid +)$$

$$\because \quad t = 0 \Rightarrow \frac{0 + 0.2}{3 + 0.2 \times 3}$$

$$\Rightarrow \frac{0 + 0.2}{3 + 0.2 \times 3} = \frac{0.2}{3.6}$$

$$\Rightarrow \boxed{P(x_2 = Y \mid +) = 0.056}$$

$$P(x_3 = G \mid +)$$

$$t = 2 \Rightarrow 2 + 0.2$$

$$\Rightarrow \underline{P(X_2 = Y \mid +)}$$

$$t = 0, \; m = 0.2, \; N = 3, \; S = 2 \{Y, N\}$$

$$\Rightarrow \frac{0 + 0.2}{3 + 0.2 \times 2} = \frac{0.2}{3.4}$$

$$\boxed{P(x_2 = Y \mid +) = 0.058}$$

$\Rightarrow$ $P(x_3 = G \mid +)$

$t = 2,\ m = 0.2,\ N = 3,$

$s = 2\ \{R, G\}$

$\Rightarrow$ $\dfrac{2 + 0.2}{3 + 0.2 \times 2} = \dfrac{2.2}{3.4}$

$$\boxed{P(x_3 = G \mid +) = 0.647}$$

$\Rightarrow$

x^1	$x2$	$x3$	
L	Y	R	−ve data
H	N	G	
M	Y	G	
H	Y	G	

$\cancel{P(x1 = L \mid -)}$

$P(x1 = L \mid -)$

$m = 0.2,\ t = 1,\ N = 4,\ S = 3\ \{L, M, H\}$

$\Rightarrow$

$$\frac{1+0.2}{4+0.2\times 3} = \frac{1.2}{4.6} = 0.26$$

$$\boxed{P(x1 = L \mid -) = 0.26}$$

$P(x2 = Y \mid -)$

$m = 0.2,\ t = 3,\ N = 4,\ s = 2\ \{Y, N\}$

$$\Rightarrow \frac{3+0.2}{4+2\times 0.2} = \frac{3.2}{4.4}$$

$$\boxed{P(x2 = Y \mid -) = 0.73}$$

$P(x_3 = G \mid -)$

$m = 0.2,\ t = 3,\ N = 4,\ s = 2\ \{R, G\}$

$$\frac{3+0.2}{4+2\times 0.2} = \frac{3.2}{4.4} = 0.73$$

$$\boxed{P(x_3 = G \mid -) = 0.73}$$

Ans 4) b)

$x = [L, Y, G]$

$P(x \mid +) = P(x = [L, Y, G] \mid +)$

$= P(L, Y, G \mid +)$

$= P(L \mid Y, G, +) \cdot P(Y, G \mid +)$

$= P(L \mid Y, G, +) \cdot P(Y \mid G, +) \cdot P(G \mid +)$

$\Rightarrow \because$ the attributes are considered as independent in the Naive Bayes, we have:

$P(L \mid Y, G, +) = P(L \mid +)$

$P(Y \mid G, +) = P(Y \mid +)$

$\Rightarrow \cancel{P(L, Y, G \mid +) = P(L \mid +) \cdot P(Y \mid +) \cdot P(G \mid}$

$\Rightarrow P(L, Y, G \mid +) = P(L \mid +) \cdot P(Y \mid +) \cdot P(G \mid +)$

$= P(x_1 = L \mid +) \cdot P(x_2 = Y \mid +) \cdot P(x_3 = G \mid +)$

$= (0.611)(0.058)(0.647)$

$= \boxed{0.023}$

$\Rightarrow$ ~~P(G.~~

$\therefore \boxed{P(L, Y, G \mid +) = 0.023}$

$\Rightarrow P(x \mid -) = P(x = [L, Y, G] \mid -)$

$= P(x_1 = L \mid -) \cdot P(x_2 = Y \mid -) \cdot P(x_3 = G \mid -)$

$= (0.26)(0.73)(0.73)$

$= 0.1386$

$\therefore \boxed{\begin{array}{l} P(x \mid -) = 0.1386 \\ P(x \mid +) = 0.023 \end{array}}$

Ans 4) c)

For ML, the label is the one which has a maximum value. Therefore,

~~P(+|+~~

$P(x|+) \quad ? \quad P(x|-)$

$0.023 \quad ? \quad 0.1386$

$0.023 < 0.1386$

$P(x|+) < P(x|-)$

$\therefore \boxed{\text{Label}_{ML} = -ve}$

Ans 4) d)

$P(+) = 3/7$

$P(-) = 4/7$

=> <u>for MAP label</u> :

$P(+|x) \quad ? \quad P(-|x)$

$\frac{P(x|+).P(+)}{P(x)} \quad ? \quad \frac{P(x|-).P(-)}{P(x)}$

$\because P(x)$ is same for both, it can be neglected

$P(x|+)\,P(+) \quad ? \quad P(x|-).P(-)$

$0.023 \times \frac{3}{7} \quad ? \quad 0.1386 \times \frac{4}{7}$

$0.069 \quad ? \quad 0.5544$

$0.069 \quad < \quad 0.5544$

$\therefore \boxed{\text{Label}_{MAP} = -ve}$