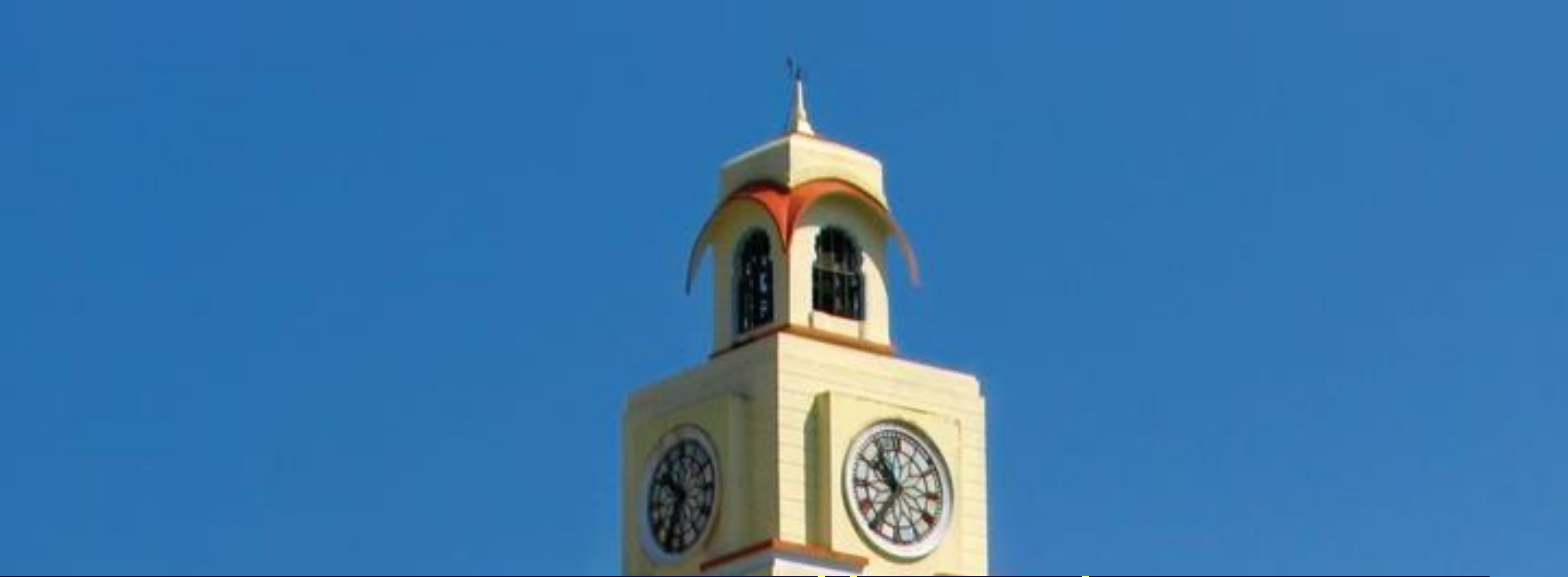

# Machine Learning
# DSECL ZG565
# Problems


Dr. Monali Mavani

BITS Pilani

Pilani Campus

# Question 1

Consider the hypothesis function $h(\mathbf{w},\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2$; with parameters

w $=< w_0 , w_1 , w_2 , w_3, w_4 > = <$ -20, -2, -4, 1, 1$>$.

Here $x_1$ and $x_2$ are two features.

- Derive the equation of the decision boundary g($x_1, x_2$) for logistic regression given by the equation:

$$y = \frac{1}{1+\exp\{-h(w,x)\}}$$

- Draw the decision boundary and predict the class labels [$C_0$, $C_1$] for the examples given by A(-2, 2), B(6, 6) and C(-5, 5).

Page No.

DATE: / /

②

$w = \langle w_0, w_1, w_2, w_3, w_4 \rangle = \langle -20, -2, -4, 1, 1 \rangle$

$h(w, x) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2$

$\sum_{i=0}^{4} w_i x_i \geq 0 \quad ; \quad x_0 = 0$

For a decision boundary

Put $\sum_{i=0}^{4} w_i x_i = 0$

$\Rightarrow \quad -20 - 2x_1 - 4x_2 + x_1^2 + x_2^2 = 0$

$x_1^2 + x_2^2 - 2x_1 - 4x_2 - 20 = 0$

$x_1^2 - 2x_1 + x_2^2 - 4x_2 = 20$

$x_1^2 - 2x_1 + 1 + x_2^2 - 4x_2 + 4 = 20 + 1 + 4$

$(x_1 - 1)^2 + (x_2 - 2)^2 = 25$

~~centre~~ centre $= (1, 2)$, $r = \sqrt{25} = 5$

[Eqn of circle $\Rightarrow (x-h)^2 + (y-k)^2 = r^2$

centre $(h, k)$, radius $= r$]

Page No.
DATE / /

$$\left[ \begin{array}{l} x^2+y^2+2gx+2fy+c=0 \\ \text{center} = (-g,-f), \quad r = \sqrt{g^2+f^2-c} \end{array} \right]$$

| | $x_1$ | $x_2$ |
|---|---|---|
| A | -2 | 2 |
| B | 6 | 6 |
| C | -5 | 5 |

If points are inside the circle then class C1

If points outside circle then class C0

(1) A (-2,2) and C(1,2) find Euclidean distance

$$d(A,C) = \sqrt{(-2-1)^2+(2-2)^2} = \sqrt{9+0} = 3$$

$3$ is $< 5$

$d(A,C) <$ radius so A is inside ~~bounda~~ circle ; $\boxed{\text{class} = C1}$

(2) B (6,6) and C (1,2).

$$d(B,C) = \sqrt{(6-1)^2+(6-2)^2} = \sqrt{25+16} = \sqrt{41}$$
$$= 6.4$$

$d(B,C) >$ radius $(6.4>5)$ so $\boxed{\text{Class} \Rightarrow C0}$

# Question 2

Consider the loss function of linear regression given by: $J(\theta_0 , \theta_1)$.
Given $(\theta_0 , \theta_1)$=0, 0.5, Estimate $\partial J/\partial\theta_1$ using the data points below:

| x | 2 | 4 | 7.0 | 8.0 | 10.0 |
|---|---|---|---|---|---|
| y | 1 | 2 | 2.5 | 3.5 | 5.5 |

DATE: / /

Linear regression $(\theta_0, \theta_1) = (0, 0.5)$

$$y = \theta_0 + \theta_1 x$$
$$y = 0.5x$$

| | | | | | |
|---|---|---|---|---|---|
| x | 2 | 4 | 7 | 8 | 10 |
| y | 1 | 2 | 2.5 | 3.5 | 5.5 |

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{n} \sum_{i=1}^{n} \left[ h_\theta(x^{(i)}) - y^{(i)} \right] x^{(i)}$$

$$= \frac{1}{5} \left[ [(0.5 \times 2) - 1] \times 2 + [(0.5 \times 4) - 2] \cdot 4 + [(0.5 \times 7) - 2.5] \cdot 7 + [(0.5 \times 8) - 3.5] \cdot 8 + [(0.5 \times 10) - 5.5] \cdot 10 \right]$$

$$= \frac{1}{5} \left[ 0 + 0 + 7 + 4 - 5 \right] = \frac{6}{5} = \underline{\underline{1.2}}$$

# Question 3

Vijay is a certified Data Scientist and he has applied for two companies -Google and Microsoft. He feels that he has a 60% chance of receiving an offer  from Google and 50% chance of receiving an offer from Microsoft. If he receives an offer from Microsoft,  he has belief that there are 80% chances of receiving an offer Google.

- What is the probability that both the companies will make an offer to him?
- If Vijay receives an offer from Microsoft, what is the probability that he will not receive an offer from Google?
- What are his chances of getting an offer from Microsoft,  considering he has an offer from Google?

Vijay is certified - - .

G => event of receiving offer from Google
M => Event " " " " Microsoft

$P(G) = 0.6$ , $P(M) = 0.5$ , $P(G|M) = 0.8$
$P(M, G)$

1) $P(G|M) = \dfrac{P(M,G)}{P(M)}$

$0.8 = \dfrac{P(M,G)}{0.5}$

$P(M,G) = 0.4$ (prob. that he receives order from both M & G)

2) prob. that he will not receive offer from google if he has received from Microsoft

$P(\bar{G}|M) = 1 - P(G|M) = 0.2$

3) prob. of getting an offer from Microsoft considering offer from Google $P(M|G)$

$P(M|G) = \dfrac{P(M,G)}{P(G)} = \dfrac{0.4}{0.6} = \underline{\underline{0.67}}$

# Question 4

Suppose that the lifetime of *Badger* brand light bulbs is modeled by an exponential distribution with (unknown) parameter $\lambda$. We test 5 bulbs and find they have lifetimes of 2, 3, 1, 3, and 4 years, respectively. What is the MLE for $\lambda$?

Light bulbs

$X_i$ = lifetime of $i^{th}$ bulb

$x_i$ = value $X_i$ takes

Given ⇒ lifetime of bulbs is modeled by an exponential distribution with unknown parameter $\lambda$

$$f_{X_i}(x_i) = \lambda e^{-\lambda x_i}$$

lifetime of bulbs are independent so joint pdf is product of individual pdf

$$f(x_1, x_2, x_3, x_4, x_5 \mid \lambda) = (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2}) \cdots (\lambda e^{-\lambda x_5})$$

$$= \lambda^n \cdot e^{-\lambda(x_1+x_2+x_3+x_4+x_5)}$$

data is fixed & $\lambda$ is variable

this density is likelihood function

$x_1 = 2$, $x_2 = 3$, $x_3 = 1$, $x_4 = 3$, $x_5 = 4$

$$f(2,3,1,3,4 \mid \lambda) = \lambda^5 \cdot e^{-13\lambda}$$

$$\ln f(2,3,1,3,4 \mid \lambda) = \ln(\lambda^5 \cdot e^{-13\lambda})$$

$$= \boxed{5\ln\lambda - 13\lambda}$$

DATE: / /

To find MLE

$$\frac{d}{d\lambda}(\log \text{liklihood}) = 0$$

$$\frac{d}{d\lambda} 5\ln\lambda - 13\lambda = 0$$

$$\Rightarrow \frac{5}{\lambda} - 13 = 0$$

$$\lambda = 5/13$$

# Question 5

The sales of a company (in million dollars) for each year are shown in the table below.

a) Find the least square regression line y = a x + b.
b) Use the least squares regression line as a model to estimate the sales of the company in 2012.

| | | | | | |
|---|---|---|---|---|---|
| x (year) | 2005 | 2006 | 2007 | 2008 | 2009 |
| y (sales) | 12 | 19 | 29 | 37 | 45 |

y = a x + b

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$b = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i\right)$$

# Question 6

Suppose we have a sample of real values, called $x_1$, $x_2$, ..., $x_n$. Each sampled from p.d.f. $p(x)$ which has the following form:

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, & if \quad x \geq 0 \\ 0, & otherwise \end{cases}$$

where $\alpha$ is an unknown parameter. Which one of the following expressions is the maximum likelihood estimation of $\alpha$? ( Assume that in our sample, all $x_i$ are large than 1. )

Ans: $\dfrac{n}{\sum_{i=1}^{n} x_i}$

Problem - 6

DATE: / /

(1)

$$f(x) = P(x) = \alpha e^{-\alpha x} \quad \text{if} \quad x \geq 0$$

$x_i$'s are larger than 1

$x_1 --- x_n$ are iid,

joint pdf is product of individual densities

$$f(x_1, x_2 --- x_n | \alpha) = (\alpha e^{-\alpha x_1})(\alpha e^{-\alpha x_2}) \cdots (\alpha e^{-\alpha x_n})$$

$$= \alpha^n e^{-\sum_{i=1}^{n} \alpha x_i}$$

Viewing data as fixed and $\alpha$ as variable

$$f(x_1 -- x_n | \alpha) = \alpha^n e^{-\sum_{i=1}^{n} \alpha x_i}$$

$$ML \Rightarrow f(x_i|\alpha) \Rightarrow \underset{\alpha}{\operatorname{argmax}} \, \alpha^n e^{-\sum_{i=1}^{n} \alpha x_i}$$

$$\log f(x_i|\alpha) = \underset{\alpha}{\operatorname{argmax}} \, \log_e \alpha^n e^{-\sum_{i=1}^{n} \alpha x_i}$$

$$= \ln \alpha^n + \ln e^{-\sum_{i=1}^{n} \alpha x_i}$$

$$\frac{d}{d\alpha} f(x_i|\alpha) = 0$$

$$\Rightarrow \frac{d}{d\alpha} \ln \alpha^n - \frac{d}{d\alpha} \sum_{i=1}^{n} \alpha x_i = 0$$

$$\Rightarrow n \cdot \frac{d}{d\alpha} \ln \alpha - \sum_{i=1}^{n} x_i \frac{d}{d\alpha} \alpha = 0$$

$$\Rightarrow \frac{n}{\alpha} - \sum_{i=1}^{n} x_i = 0$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} x_i \quad ; \quad \boxed{\alpha = \frac{n}{\sum_{i=1}^{n} x_i}}$$

# Question 7

- Derive the maximum likelihood estimator (MLE) for the mean μ of a univariate normal distribution. Assume N samples, $x_1, \ldots, x_N$ independently drawn from a normal distribution with known variance $\sigma^2$ and unknown mean μ. Show all intermediate steps and assumptions.

# Question 8

- Given $N$ independent measurements $x_1, x_2, \dots, x_N$, determine the optimal parameters of the model, i.e. the parameters that maximize the probability density function (PDF). To model this data, assume Gaussian distribution.

# Question 9

- Consider a dataset for binary classification problem with class labels [$C_1$, $C_0$]. The features are given by $F_1$, $F_2$ and $F_3$. Each of these features have two values as given in the dataset below. Apply Naïve Bayes classifier by computing the probabilities to classify the new example: <$F_1$=$x_1$, $F_2$=$y_2$, $F_3$=$z_1$>

| Sl No | $F_1$ | $F_2$ | $F_3$ | Class |
|---|---|---|---|---|
| 1 | $x_1$ | y2 | z1 | $C_1$ |
| 2 | x2 | $y_1$ | z2 | $C_0$ |
| 3 | $x_1$ | $y_1$ | z2 | $C_1$ |
| 4 | x2 | y2 | z1 | $C_0$ |
| 5 | x2 | $y_1$ | z1 | C1 |
| 6 | x2 | y2 | z1 | $C_0$ |
| 7 | $x_1$ | $y_1$ | z2 | $C_1$ |
| 8 | $x_1$ | y2 | z2 | $C_0$ |

DATE: / /

(5) features => $F_1, F_2, F_3$
classes => $C_0, C_1$
test => $(x_1, y_2, z_1)$

(1) $P(C_0) = \frac{\text{no. of records with } C_0}{\text{Total examples}}$
$= \frac{4}{8} = 0.5$

(2) $P(C_1) = 4/8 = 0.5$

(3) $P(\overset{x_1}{F_1}|C_0) = \frac{\text{no. of records with } F_1 \overset{=x_1}{} \text{ \& } C_0}{\text{no. of records with } C_0}$
$= \frac{1}{4}$

(4) $P(y_2|C_0) = \frac{3}{4}$

(5) $P(z_1|C_0) = 2/4$

Naive Bayes =>
(I) $P(y=C_0) \prod_i P(x^{Test_i}|y=C_0)$
$P(y=C_1) \prod_i P(x^{Test}|y=C_1)$

(I) $P(C_0) . P(x_1|C_0) . P(y_2|C_0) . P(z_1|C_0) = (\frac{1}{4})(\frac{3}{4})(\frac{2}{4})(\frac{4}{8})$
$= 3/64$

II $P(C_1) . P(x_1|C_1) . P(y_2|C_1) P(z_1|C_1) = 3/64$
both are same so pick any

Thank you