Q.1. We have five data points $x_1 = 1, x_2 = 3, x_3 = -1, x_4 = 4, x_5 = -3$ which are obtained from sampling a Gaussian distribution of zero mean. What is the Maximum Likelihood Estimate of the variance of the Gaussian distribution? Show all the steps in the calculation. [5 Marks]

We are sampling f a normal distribution $N(0, \sigma^2)$ where σ is

The likelihood is $\prod_{i=1}^{5} P(x_i / N(0, \sigma^2))$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2/\sigma^2}\right)\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2/\sigma^2}\right) \cdots \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{x_5^2}{\sigma^2}}\right)$$

$$L = \frac{1}{(2\pi)^{5/2}\sigma^5} e^{-\frac{1}{2}\left(\frac{x_1^2 + x_2^2 + \cdots x_5^2}{\sigma^2}\right)}$$

To find the value of that maximizes likelihood we have to ca the derivative with respect to σ of L and equate it to 0.

It is more convenient to work with $\log L$

$$\frac{d}{d\sigma}(\log L) = 0$$

$$\Rightarrow \frac{d}{d\sigma}\left(\log\frac{1}{(2\pi)^{5/2}} + \log\frac{1}{\sigma^5} + -\frac{1}{2}\left(\frac{x_1^2 + x_2^2 + \cdots x_5^2}{\sigma^2}\right)\right) = 0$$

$$\Rightarrow \frac{d}{d\sigma}\left(-5\log\sigma - \frac{x_1^2 + x_2^2 + \cdots x_5^2}{2\sigma^2}\right) = 0$$

$$\Rightarrow \frac{-5}{\sigma} + \frac{1}{\sigma^3}\left(x_1^2 + x_2^2 + \cdots x_5^2\right) = 0$$

Marking Scheme
2 Marks → Setting up max likelihood expression
3 Marks - differentiation and final calculation

Q.2. Consider a table with a single attribute "wind" and category "rain", where "wind" can take two attribute values – high and low, and "rain" has two classes – yes and no. There are 10 entries in the table, and it is known that 8 entries in the table have wind=high. It is also known that 8 entries in the table also have rain=yes. What is the highest and lowest possible information gain if we split the table on the attribute "wind"? [5 Marks]

For the highest information gain the table looks like this:

Wind	Rain
High	Yes
High	Yes
High	Yes
High	Yes
High	Yes
High	Yes
High	Yes
High	Yes
Low	No
Low	No

$$\text{Gain}(S, \text{Wind}) = \text{Entropy}(S) - \frac{|S_{w=H}|}{|S|}\text{Entropy}(S_{w=H}) - \frac{|S_{w=L}|}{|S|}\text{Entropy}(S_{w=L})$$

$$Entropy(S) = -p_+ \log p_+ - p_- \log p_- =$$

$$= -\frac{8}{10}\log\frac{8}{10} - \frac{2}{10}\log\frac{2}{10} = 0.721$$

$$Entropy(S_{w=H}) = -\frac{8}{8}\log\frac{8}{8} - \frac{0}{8}\log 0 = 0$$

$$Entropy(S_{w=L}) = \frac{0}{2}\log\frac{0}{2} - \frac{2}{2}\log 1 = 0$$

Highest Information gain $= 0.721 - \frac{8}{10}\times 0 - \frac{2}{10}\times 0 = \underline{0.721}$

For the lowest information gain the table looks like this:

Wind	Rain
High	No
High	No
High	Yes
High	Yes
High	Yes
High	Yes
High	Yes
High	Yes
Low	Yes
Low	Yes

In this case

$$Gain(S, wind) = Entropy(S) - \frac{|S_{w=H}|}{|S|} Entropy(S_{w=H})$$

$$- \frac{|S_{w=L}|}{|S|} Entropy(S_{w=L})$$

$$\text{Entropy}(S_{w=H}) = -\frac{6}{8}\log\frac{6}{8} - \frac{2}{8}\log\frac{2}{8} = 0.8075$$

$$\text{Entropy}(S_{w=L}) = -\frac{2}{2}\log\frac{2}{2} - \frac{0}{2}\log\frac{0}{2} = 0$$

$$\text{Gain}(S, \text{Wind}) = 0.721 - \frac{8}{10}\times 0.8075 - \frac{2}{10}\times 0$$
$$= \underline{0.075}$$

Marking Scheme

Highest Information Gain = 2.5 marks

Lowest Information Gain = 2.5 marks

Q.3. What is the best curve of the form $y = a + bx + cx^2$ in terms of minimizing square error that fits the following data of the form (x, y): (-1,0), (1,10), (2,24), (-2,4)? [5 Marks]

Computing the square loss function gives

$$L = (a - b + c - 0)^2 + (a + b + c - 10)^2 + (a + 2b + 4c - 24)^2 + (a - 2b + 4c - 4)^2$$

Setting $\frac{dL}{da} = 0$, $\frac{dL}{db} = 0$ and $\frac{dL}{dc} = 0$

We get

$$4a + 10c = 38$$

$$4a + 16b + 4c = 100$$

$$20a + 68c = 244$$

Final answer: $a = 2$, $b = 5$, $c = 3$

Marking Scheme

Setting up loss function = 2 marks

Final calculation = 3 marks

Q.4. There exists a training set consisting of 100 documents for text classification consisting of two types of document '+' and '-'. 75 of the 100 documents are '+' and the remaining are '-'. The total number of words including duplicates in the '+' documents is 150 and the total number of words including duplicates in the '-' documents is 100. The number of words in the vocabulary is 1000. What classification is given to a test text with 5 words consisting of words belonging to the vocabulary but which have not occurred in the training set at all using the Naïve Bayes algorithm for text classification?

[5 Marks]

We have $P(+) = \frac{75}{100} = \frac{3}{4}$

$P(-) = \frac{25}{100} = \frac{1}{4}$

Since the words ω_K in the test document do not occur in the training set, we have

$$P\left(\frac{\omega_K}{+}\right) = \frac{n_K + 1}{n + |V|} = \frac{0+1}{150+1000} = \frac{1}{1150}$$

$$P\left(\frac{\omega_K}{-}\right) = \frac{n_K + 1}{n + |V|} = \frac{1}{100+1000} = \frac{1}{1100}$$

$$P(+)\prod_{K=1}^{5} P\left(\frac{\omega_K}{+}\right) = \frac{3}{4}\left(\frac{1}{1150}\right)^5$$

$$P(-)\prod_{K=1}^{5} P\left(\frac{\omega_K}{-}\right) = \frac{1}{4}\left(\frac{1}{1100}\right)^5$$

$$\frac{3}{4}\left(\frac{1}{1150}\right)^5 > \frac{1}{4}\left(\frac{1}{1100}\right)^5$$

$\therefore$ we give a +ve classification to the new text.

Marking Scheme

3 marks $\rightarrow$ setting up $P\left(\frac{wk}{+}\right)$ and $P\left(\frac{wk}{-}\right)$

2 marks $\rightarrow$ final calculation

Q.5. One percent of women over 50 have breast cancer. Ninety percent of women who have breast cancer test positive on mammograms. Eight percent of women will have false positives. What is the probability that a woman has cancer if she has a positive mammogram result? [5 Marks]

Let C denote the event of cancer for women over 50.

$$P(C) = 0.01$$

$$P(M/C) = 0.9$$

$P(M/\neg C)$ = probability of showing a positive mammogram test given no cancer = false positive = 0.08

$P(C/M) = ?$

$$P(C/M) = \frac{P(M/C)\,P(C)}{P(M/\neg C)\,P(\neg C) + P(M/C)\,P(C)}$$

$$= \frac{(0.9)(0.01)}{(0.08)(0.99) + (0.9)(0.01)}$$

$$= \frac{0.009}{0.0882} = 0.102$$

<u>Marking scheme</u>:

2 Marks → calculating all the probabilities $P(C)$, $P(\neg C)$, $P(M/C)$ $P(M/\neg C)$

3 Marks → Bayes equation

Q.6. Let X_1, X_2 be two real-valued features and Y be a Boolean-valued function of the given features such that the Gaussian Naïve-Bayes assumptions are satisfied. Suppose $P(Y/X_1, X_2) = \frac{1}{1+\exp(-0.1-0.2X_2-0.3X_3)}$. Assume that $P(X_1/Y=0) = N(1.0, \sigma_1)$ and $P(X_1/Y=1) = N(2.0, \sigma_1)$. Similarly $P(X_2/Y=0) = N(1.0, \sigma_2)$ and $P(X_2/Y=1) = N(2.0, \sigma_2)$ Calculate the standard deviations σ_1 and σ_2 and the probability $P(Y=1)$. [5 Marks]

From the formula for logistic regression we see that

$$\frac{\mu_{10} - \mu_{11}}{\sigma_1} = \text{coefficient of } X_i \text{ in the}$$

$$\text{formula } P(Y=1/X) = \frac{1}{1+\exp(\theta_0 + \Sigma \theta_i X_i)}$$

$$\therefore \frac{1.0 - 2.0}{\sigma_1^2} = -0.2 \Rightarrow \sigma_1^2 = 5$$

$$\frac{1.0 - 2.0}{\sigma_2^2} = -0.3 \Rightarrow \sigma_2^2 = \frac{10}{3}$$

$$\text{Also } \ln\left(\frac{1-\pi}{\pi}\right) + \frac{\mu_{11}^2 - \mu_{10}^2}{2\sigma_1^2} + \frac{\mu_{21}^2 - \mu_{20}^2}{2\sigma_2^2}$$

$= \text{constant term where } \pi = \text{prior prob}$
$= P(Y=1)$

$$\ln\frac{1-\pi}{\pi} + \frac{4-1}{2\times 5} + \frac{4-1}{2\times\frac{10}{3}} = -0.1$$

$$\ln\frac{1-\pi}{\pi} = -0.1 - \frac{3.0}{10} - \frac{9}{20} = -0.85$$

$$\frac{1-\pi}{\pi} = e^{-0.85}$$

$$\frac{1}{\pi} = 1 + e^{-0.85}$$

$$\Rightarrow \pi = \frac{1}{1+e^{-0.85}}$$

$$\frac{\mu_{i0} - \mu_{i1}}{\sigma_i} = \text{Coefficient of } X_i \text{ in the}$$

formula $P(Y=1/X_1, X_2) = \dfrac{1}{1+\exp(\theta_0 + \Sigma\theta_i X_i)}$

Marking Scheme

2 Marks $\rightarrow$ getting the formulae right

3 Marks $\rightarrow$ final calculation