Problem 1 :-

P(A | B) means Probability of A, given B.
P(B | A) means Probability of B, given A.

Lets say A is an event where a person ~~takes~~ goes out ~~go~~.
Lets say B is an event where it rains.
∴ P(A|B) means ~~Probability~~ Probability of a person going out given that it rains.
which is true, we all check the weather getting out.

Now, P(B|A) means Probability ~~o~~ of an event where it rains, given that a person goes out.
This isn't true because we all know that it doesn't rain after checking if someone is going out.

Hence, we can say that P(A|B) and P(B|A) are in general not the same.

Problem 2 :-

1.) X and Y are independent,

$\therefore E(X, Y) = E(X)E(Y)$

$\therefore Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y)$

$= 0$

$\therefore \rho_{xy} = \dfrac{Cov(X, Y)}{\sigma_x \sigma_y} = \dfrac{0}{\sigma_x \sigma_y} = 0$

$\because$ the correlation is 0, the variables X & Y are uncorrelated.

2.) Its given that X & Y are uncorrelated

Lets consider X to be uniform distribution over $[-1, 1]$ &

$Y = X^2$

$E[XY | X \leq 0] = \int_{-1}^{0} -x^2 \, dx = -1/3$

$E[XY | X > 0] = \int_{0}^{1} x^2 \, dx = 1/3$

$E[XY] = 0$

$\therefore$ Since the joint distribution of X & Y is not uniform, we can say that they are not independent.

Problem 3 :-

Given :- Components of vector $x = [x_1, x_2, \dots, x_d]^T$ are binary values (0 or 1).

$p_{ij} = P(x_i = 1 | \omega_j)$, $i = 1, \dots, d$, $j = 1, \dots, c$.

$p_{ij} = P(x_i = 1 | \omega_j)$

∴ It is the probability of $x_i = 1$ when ω_j is given.

Let consider the following discriminant function :-

$$g_j(x) = \ln[P(\frac{x}{\omega_j}) P(\omega_j)] = \ln(\frac{x}{\omega_j}) + \ln P(\omega_j)$$

Since the components of x are statistically independent for x at ω_j, we can write the density as;

$$P(x|\omega_j) = \prod_{i=1}^{d} P(x_i | \omega_j)$$

$$= \prod_{i=1}^{d} P_{ij}^{x_i} (1 - P_{ij})^{1 - x_i}$$

Thus we have the discriminant function.

$$g_j(x) = \sum_{i=1}^{d} [x \ln P_{ij} + (1 - x_i) \ln (1 - P_{ij})] + \ln P(\omega_j)$$

$$\therefore g_j(x) = \sum_{i=1}^{d} x_i \ln \frac{P_{ij}}{1 - P_{ij}} + \sum_{i=1}^{d} \ln(1 - P_{ij}) + \ln P(\omega_j)$$

Problem 4 :-

Consider 2-class classification with feature vector x, suppose $p(x|\omega_1)$ is standard normal distribution and $p(x|\omega_2)$ is uniform distribution over $[-\frac{1}{2}, \frac{1}{2}]$

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$p(x|\omega_2) = 1, \quad x \in [-\frac{1}{2}, \frac{1}{2}]$$

Assuming ~~zero loss~~ ~~&~~ zero-one loss & $P(\omega_1) = P(\omega_2)$, using likelihood ratio test to derive the corresponding decision rule.

If $x \in [-\frac{1}{2}, \frac{1}{2}]$, then

$$\cancel{p(x|x_1)}\; p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \leq 1 = p(x|\omega_2)$$

$$\therefore \frac{P(x|\omega_1)\,P(\omega_1)}{P(x)} \leq \frac{P(x|\omega_2)\,P(\omega_2)}{P(x)}$$

$$\cancel{P(x|\omega_1)P(\omega_1) \leq P(x|\omega_2)P(\omega_2)}$$

$$\therefore P(\omega_1|x) \leq P(\omega_2|x)$$

Lets choose class (ω_2)

If $x \notin [-\frac{1}{2}, \frac{1}{2}]$, then

$$p(x|\omega_1) \geq 0 = p(x|\omega_2)$$

$$\therefore P(\omega_1|x) \geq P(\omega_2|x)$$

Lets choose class (ω_1)

$x \in \omega_1, \omega_2 \Rightarrow [-\ ,\ ]$

$p(\omega_1) = \quad \leq \quad = p(x|\omega_2)$

$$\therefore \frac{P(x|\omega_1)\,P(\omega_2)}{P(\quad)} \leq \frac{P(x|\omega_2)\,P(\omega_2)}{P(\quad)}$$

Lets choose class ω_2

$x \in [- \quad , \quad]$, then

$p(x|\omega_1) \geq 0 = p(x|\omega_2)$

$\therefore P(\omega_1|x) \geq P(\omega_2|x)$

$\therefore P(\omega_1|x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$P(\quad) = 1, \; x \in [-\frac{1}{2}, \frac{1}{2}]$

Problem 5 :-

We have been given:-

$$\lambda\left(\frac{\alpha_i}{\omega_j}\right) = \begin{cases} 0 & i=j \quad i,j=1,\dots,c \\ \lambda_r & i=c+1 \\ \lambda_s & \text{otherwise} \end{cases}$$

We have c ~~pos~~ classes $\{\omega_1, \omega_2, \dots, \omega_c\}$

We have a possible actions $\{\alpha_1, \alpha_2, \dots, \alpha_a\}$

The loss function is the loss incurred for taking action α_i when the class is ω_j; i.e. $\lambda\left(\frac{\alpha_i}{\omega_j}\right)$

We can consider the loss that would incur due to each possible actions.

Now, expected loss or the conditional risk is :-

$$R\left(\frac{\alpha_i}{x}\right) = \sum_{j=1}^{c} \lambda\left(\frac{\alpha_i}{\omega_j}\right) P\left(\frac{\omega_j}{x}\right)$$

The zero-one conditional risk is :-

$$R\left(\frac{\alpha_i}{x}\right) = \sum_{j \neq 1} P\left(\frac{\omega_j}{x}\right)$$

$$= 1 - P\left(\frac{\omega_i}{x}\right)$$

1) For $i = 1, 2, \dots, c$.

$$R\left(\frac{\alpha_i}{x}\right) = \sum_{j=1}^{c} \lambda\left(\frac{\alpha_i}{\omega_j}\right) P\left(\frac{\omega_j}{x}\right)$$

$$= \lambda_s \sum_{j=1, j\neq 1}^{c} P\left(\frac{\omega_j}{x}\right)$$

Using zero-one conditional risk, we get

$$= \lambda_s \left[1 - P\left(\frac{\omega_j}{x}\right)\right]$$

For $i = c+1$;

$$R\left(\frac{\alpha_{c+1}}{x}\right) = \lambda_r$$

In order to make the risk minimum, we decide if $R\left(\frac{\alpha_1}{x}\right) \leq R\left(\frac{\alpha_{c+1}}{x}\right)$

$$P\left(\frac{\omega_i}{x}\right) \geq 1 - \frac{\lambda_r}{\lambda_s}$$ and reject otherwise

2.) If $\lambda_r = 0$, then

$$P\left(\frac{\omega_i}{x}\right) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

if $\lambda_r = 0$, then the probability will be greater than 1 which isn't possible, hence we'll reject.

3.) If $\lambda_r > \lambda_s$, then

$$P\left(\frac{\omega_1}{x}\right) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

if $\lambda_r > \lambda_s$, we will never reject.

Problem 6:-

Probability density general representation:-

$$p(x|\eta) = h(x)\exp\{\eta^T T(x) - A(\eta)\}$$

Where η is natural parameter

$h(x)$ = the base density which ensures x is in right space.

$T(x)$ = the sufficient statistics

$A(\eta)$ = the log normalizer which is determined by $T(x)$ & $h(x)$

$\exp(\cdot)$ represents the exponential function.

1.) We know that, $\int_{-\infty}^{\infty} p(x|\eta)\,dx = 1$

$$\int h(x)\exp\{\eta^T T(x) - A(\eta)\}dx = 1$$

$$\therefore \exp\{-A(\eta)\}\int h(x)\exp\{\eta^T T(x)\}dx = 1 \quad \text{--- ①}$$

$$\therefore -A(\eta) = -\ln\left\{\frac{1}{\int h(x)\exp\{\eta^T T(x)\}dx}\right\}$$

$$A(\eta) = \ln\{\int h(x)\exp\{\eta^T T(x)\}dx\}$$

2.) $$\frac{\partial}{\partial\eta}A(\eta) = \frac{1}{\int h(x)\exp\{\eta^T T(x)\}dx} \cdot \int h(x)\frac{\partial}{\partial\eta}\exp\{\eta^T T(x)\}dx$$

Using ①

$$= \exp\{-A(\eta)\}\int h(x)T(x)\exp\{\eta^T T(x)\}dx$$

$\left[\because \frac{\partial}{\partial A} cA^T = c,\right]$

$= \int T(x) h(x) \exp\{\eta^T T(x) - A(\eta)\} dx$

$= \int T(x) p(x|\eta) dx = E_\eta(T(x))$

3.) Likelihood of η = $L(\eta) = \prod_{i=1}^{n} p(x_i|\eta)$

$\ln L(\eta) = \ln\left(\prod_{i=1}^{n} p(x_i|\eta)\right)$

$= \ln\left(\prod_{i=1}^{n} (h(x_i) \exp\{\eta^T T(x_i) - A(\eta)\}\right)$

We need to maximize $\ln L(\eta)$ w.r.t η to find M.L.E

$\because \frac{\partial}{\partial \eta} \ln L(x) \geq 0$

$\frac{\partial}{\partial \eta} \ln(\exp\{\eta^T \sum T(x_i) - nA(\eta)\}) \geq 0$

$\frac{\partial}{\partial \eta} \left\{\eta^T \sum_{i=1}^{n} T(x_i) - nA(\eta)\right\} \geq 0$

$\sum_{i=1}^{n} T(x_i) - nE_\eta(T(x)) \geq$

$E_\eta(T(x)) \geq \frac{1}{n} \sum_{i=1}^{n} T(x_i) \quad \therefore E_\eta(T(x)) = \frac{1}{n} \sum_{i=1}^{n} T(x_i)$

$\times$