Neil. Gupte.
10445674.

Cs-559 HomeWork 1

1 a) Having zero covariance is same as being uncorrelated.

$$Cov[x,y] = E[xy] - E[x]E[y]$$

$$\therefore E[xy] = E[x]E[y] \quad \because cov[x,y]=0$$

If X and Y are independent then →.

$$E[xy] = \int\int xy\, p_{x,y}(x,y)\,dx\,dy$$

$$= \int\int xy p_x(x)\; p_y(y) dx dy$$

$$= \int x p_x(x) \left(\int y\, p_y(y) dy\right) dx$$

$$= \left(\int x p_x(x) dx\right)\left(\int y\, p_y(y) dy\right)$$

$$= E[x]\; E[y]$$

b). Suppose X and X are uncorrelated, then we cannot conclude X and Y are independent. They can be dependent.
For example →.
$X \sim N(0,1)$ $\quad Y = X^2$ $\quad$ Y is dependent on X

1] 1) $E(X) = 0$ 2) $E(X)^3 = 0$ for ND.

$$\operatorname{cov}(x,y) = E(XY) - E(X)E(Y)$$

$$= E(X X^2) - E(X)E(Y).$$

$$= E(X^3) - E(X)E(Y]$$

$$= 0 - 0 \cdot E(Y).$$ from 1) and 2)

$$\operatorname{cov}(x,y) = 0.$$

$$\rho(x,y) = \frac{\operatorname{cov}[x,y]}{\sigma_x \cdot \sigma_y} = 0.$$

Hence proved though x, y are uncorrelated they are dependent.

2] 1) Since $P(\omega_{max}|x) \geq P(\omega_i|x)$ we have.

$$\sum_{i=1}^{c} P(\omega_{max}|x) \geq \sum_{i=1}^{c} P(\omega_i|x) = 1.$$

Hence

$$c\,P(\omega_{max}|x) \geq 1$$

implies.

$$P(\omega_{max}|x) \geq \frac{1}{c}.$$

2] b)

$$P(\text{error}) = \int_{\Omega} P(\text{error}|x)p(x)dx$$

$$= \int_{\Omega} [1 - P(\omega_{max}|x)]p(x)dx$$

~~$=\int_{\Omega}$~~

$$= 1 - \int_{\Omega} P(\omega_{max}|x)p(x)dx.$$

c) From a) and b)

$$P(\text{error}) = 1 - \int_{\Omega} P(\omega_{max}|x)p(x)dx$$

$$\leq 1 - \int_{\Omega} \frac{1}{C} p(x)dx$$

$$\leq 1 - \frac{1}{C}. \quad \leq \frac{C-1}{C}.$$

3] Given → $P(x|\omega_1) = N(4,1)$ $\mu=4$ $\sigma=1$.
$P(x|\omega_2) = N(8,1)$ $u=8$ $\sigma=1$.
prior probability =

$P(\omega_2) = \frac{1}{4}$. $\quad P(\omega_1) = 1 - \frac{1}{4} = \frac{3}{4}.$

$$\lambda = \begin{bmatrix} \overset{\lambda_{11}}{0} & \overset{\lambda_{12}}{3} \\ \overset{\lambda_{21}}{1} & \overset{\lambda_{22}}{0} \end{bmatrix}$$

By likelihood ratio test ⇒.

$$\frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)} .$$

LHS =

$$\frac{\frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-4}{2}\right)^2}}{\frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-8}{2}\right)^2}} > e^{-\left(\frac{x-4}{2}\right)^2+\left(\frac{x-8}{2}\right)^2}$$

RHS =

$$\frac{3-0}{1-0} \frac{P(\omega_2)}{P(\omega_1)} > \frac{\cancel{3}}{1}\left(\frac{1}{\cancel{4}} \times \frac{\cancel{4}}{\cancel{3}}\right) = 1.$$

Combining both LHS and RHS.

$$e^{-\left(\frac{x-4}{2}\right)^2+\left(\frac{x-8}{2}\right)^2} > 1.$$

Taking logs ⇒

$$-\left(\frac{x-4}{2}\right)+\left(\frac{x-8}{2}\right)^2 \ln e > \ln 1 .$$

$$\frac{-x^2+8x-16+x^2-16x+64}{4} > 0 .$$

$$\frac{-8x+48}{4} > 0$$

$$-2x+12 > 0$$

$$-2x > -12 .$$

∴ If $x < 6$, decide ω_1

$x > 6$ decide ω_2.

4]

a). For $i = 1, \ldots\ldots, c,$

$$R(\alpha_1|x) = \sum_{j=1}^{c} \lambda(\alpha_i|w_j)\, P(w_j|x).$$

$$= \lambda_s \sum_{j=1, j\neq i}^{c} P(w_j|x).$$

$$= \lambda_s [1 - P(w_i|x)].$$

For $i = c+1$.

$$R(\alpha_{c+1}|x) = \lambda r$$

Therefore, minimum risk if we decide w_i if $R(\alpha_i|x) \leq R(\alpha_{c+1}|x)$

i.e $P(w_i|x) \geq 1 - \frac{\lambda r}{\lambda s}$ reject

otherswise,

b). If $\lambda r = 0$, we will always reject.

c) If $\lambda_r > \lambda_s$, we will never reject.

5] i) The likelihood Function is →.

$$L(\theta|x) = f(x_1|\theta) \ldots f(x_n|\theta).$$

$$= \theta e^{-\theta x_1} \ldots \theta e^{-\theta x_n}$$

$$= \theta^n e^{-\theta(x_1 \ldots + \ldots x_n)}.$$

Taking log →.

$$\ln L(\theta|x) = n \ln \theta - \theta(x_1 + \ldots x_n)$$

~~Its log~~ Gradient wrt θ.

$$\frac{d}{d\theta} \ln L(\theta|x) = \frac{n}{\theta} - (x_1 + \ldots x_n)$$

Comparing it to 0 →.

$$\frac{n}{\theta} = x_1 + \ldots x_n$$

$$\theta = \frac{n}{x_1 \ldots x_n}$$

MLE ∴ $\hat{\theta} = \frac{1}{\bar{x}}$

5] b) $f_{X_1}(x_1 ; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_1-\theta_1)^2}{2\theta_2}}$

Likelihood function.

$L(x_1 \ldots x_n ; \theta_1, \theta_2)$ ~~$= -n \ln(2\pi)$~~ ~~2~~

$$= \frac{1}{2\pi^{\frac{n}{2}} \theta_2^{\frac{n}{2}}} \exp\left(-\frac{1}{2\theta_2} \sum_{i=1}^{n} (x_1-\theta_1)^2\right)$$

Taking LLD →.

$$\ln L(x_1 \ldots x_n ; \theta_1 \theta_2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \theta_2 - \frac{1}{2\theta_2} \sum_{I=1}^{n} (x_1 - \theta_1)^2.$$

Taking gradients wrt θ_1 and θ_2.

$$\frac{\partial}{\partial \theta_1} \ln L(x_1 \ldots \theta_1, \theta_2) = \frac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1) = 0$$

$$\frac{\partial}{\partial \theta_2} \ln L(x_1 \ldots \theta_1, \theta_2) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_1-\theta_1)^2 = 0$$

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

6] a) As there are only two output classes → $y = 0$ or $y = 1$ we can use sigmoid function for classification.

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

$$x^T\theta = \theta^T x = \sum_{i=1}^{m} \theta_i x_i$$

$$P(Y = y \mid X = x) = \sigma(\theta^T x)^y \cdot \left[1 - \sigma(\theta^T x)\right]^{(1-y)}.$$

(Bernoulli Princi).

$$L(\theta) = \prod_{i=1}^{n} \sigma(\theta^T x)^y \cdot \left[1 - \sigma(\theta^T x^{(i)})\right]^{(1-y^{(i)})}$$

Taking log.

$$LL(\theta) = \sum_{i=1}^{n} y^{(i)} \log \sigma(\theta^T x^{(i)}) + (1 - y^{(i)}) \cdot \log\left[1 - \sigma(\theta^T x^{(i)})\right].$$

= Taking gradient wrt θ.

$$\frac{\partial LL(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} y \log \sigma(\theta^T x) + \frac{\partial}{\partial \theta_j} (1-y) \log\left[1 - \sigma(\theta^T x\right]$$

$$= \left[\frac{y}{\sigma(\theta^T x)} - \frac{1-y}{1 - \sigma(\theta^T x}\right] \frac{\partial}{\partial \theta_j} \sigma(\theta^T x).$$

$$= \left[\frac{y}{\sigma(\theta^T x)} - \frac{1-y}{1 - \sigma(\theta^T x)}\right] \sigma(\theta^T x] \left[1 - \sigma(\theta^T x)\right] x_j$$

After cross multiplying:

$$= \left[y - \sigma(\theta^T x)\right] x_j$$