SML - Assignment 3

-Harkishan Singh (2017233)

Harkishan Singh (2017233) Date ....

SML - Assignment-3

Q.1 $$P(x|\theta) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i}$$

$\theta = (\theta_1, --, \theta_d)^T$

To show : $$\hat{\theta} = \frac{1}{n} \sum_{k=1}^{n} x_k$$

Consider the distribution of multivariate Bernoulli distribution :-

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

Its likelihood function (L) will be : $\prod_{k=1}^{n} P(x_k|\theta)$

so, $$L = \prod_{k=1}^{n} P(x|\theta) = \prod_{k=1}^{n} \prod_{i=1}^{d} \theta_i^{x_i^k} (1-\theta_i)^{1-x_i^k}$$

(Note: $x_i^k \Rightarrow$ i is for indexing of inner loop and k is for indexing outer loop).

Take Log on both sides

$$\log(L) = \log\left[\prod_{k=1}^{n} \prod_{i=1}^{d} \theta_i^{x_i^k}(1-\theta_i)^{1-x_i^k}\right]$$

Next page $\Rightarrow$

Date ....

$$\log(L) = \sum_{k=1}^{n} \sum_{i=1}^{d} \left[x_i^k \log(\theta_i) + (1-x_i^k)\log(1-\theta_i)\right]$$

$[\because$ log will change Π to $\Sigma]$

Differenciate above eq. wrt θ

$$\frac{\partial \log L}{\partial \theta} = \sum_{k=1}^{n} \left[\frac{x^k}{\theta} - \frac{(1-x^k)}{1-\theta}\right] = 0$$

$\left(\sum_{i=1}^{d}\right.$ is dropped off $\Downarrow$ summed up $\left.\right)$

$\theta = (\theta_1, \theta_2, --, \theta_n)$

$$\frac{\sum_{k=1}^{n} x^k}{\theta} - \frac{n - \Sigma x^k}{1-\theta} = 0 \Rightarrow \frac{\Sigma x^k}{\theta} = \frac{n - \Sigma x^k}{1-\theta}$$

$$\frac{1}{\theta} - \cancel{1} = \frac{n}{\Sigma x^k} - \cancel{1} \Rightarrow \boxed{\hat{\theta}_{ML} = \frac{1}{n}\sum_{k=1}^{n} x^k}$$

Q.2

$p(x|\omega_1) \sim N(0,1)$

$p(x|\omega_2) \sim N(\mu,1)$ & $p(x|\omega_2) \sim N(1,10^6)$ $\downarrow$ poor model

$p(\omega_1) = p(\omega_2) = 1/2$

a) To get estimate of μ in poor model, we know that for gaussian model :-

$$\mu = \bar{X} = \frac{1}{n}\Sigma x_i$$

given large data we will have $\boxed{\mu_{ML} = 1}$

b) Decision Boundary (DB) :- for $p(x|\omega_1) \sim N(0,1)$ and $p(x|\omega_2) \sim N(1,1)$

$p(\omega_1) = p(\omega_2) = 1/2$

next page

Date ....

DB: $\frac{1}{2}\frac{1}{\sqrt{2\pi 1^2}}\exp\left[-\frac{(x-0)^2}{1^2}\right] = \frac{1}{2}\frac{1}{\sqrt{2\pi 1^2}}\exp\left[-\frac{(x-1)^2}{1}\right]$

x: DB

$\boxed{DB = x = \frac{1}{2}}$ So, we have DB = 0.5 for the specified 2 classes.

c) Assumption/Note: In the question we are taking about true distribution, but we are provided the distribution of assumed class (i.e. $N(\mu, 1)$) for which we have already computed DB in b) part. So, I am assuming distribution to be $N(1, 10^6)$

Two classes we have are :-

$p(x|\omega_1) \sim N(0,1)$ $\quad$ $p(x|\omega_2) \sim N(1, 10^6)$

$p(\omega_1) = p(\omega_2) = \frac{1}{2}$

for DB :-

DB: $\frac{P(x|\omega_1) \cdot P(\omega_1)}{P(x)} = \frac{P(x|\omega_2)\, P(\omega_2)}{P(x)}$

$$\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(x-0)^2}{1}\right] = \frac{1}{\sqrt{2\pi 10^6}}\exp\left[-\frac{(x-1)^2}{10^6}\right]$$

Take log :-

$$-x^2 = -3\ln 10 - \frac{(x-1)^2}{10^6}$$

$$-10^6 x^2 = -3\times 10^6 \ln 10 - (x^2+1-2x)$$

$$x^2(10^6-1) + 2x - (3\times 10^6 \ln 10 + 1) = 0$$

Solving the above quadratic equation gives :-

DB: $x = 8.311, -8.311$

So, b/w these 2 points, we choose ω_1 else ω_2 :

$$\text{DB} = \text{Decision} \begin{cases} \omega_1 & \text{if } x \in [-8.311, 8.311] \\ \omega_2 & \text{o/w.} \end{cases}$$

Spiral

Date ....

Q.3 $D = x_1, x_2, \ldots, x_n$ $\mu' = x_1$

The true mean of data D is given by :-

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

μ : True mean

a) To show :- $\mu' = x_1$ is unbaised

$\mu' = x_1$ To show that $\mu' = x_1$ is unbaised, we need to show that $E(\mu')$ is equal to the true mean of data. So,

$$\boxed{E(\mu') = E(x_1) = \underbrace{\mu}_{\text{True mean of data}}}$$

[$\because$ expectation of any data point = mean of data]

So, it is unbaised

b) It is undesirable :-

Consider the data : $D = 10^6, -10^6$ with 2 data points

True mean of $D = \mu = \frac{1}{2}(10^6 - 10^6) = 0$

But the mean from above part $\mu' = 10^6 = x_1$

If we compute mean absolute error :-

$e = |\mu - \mu'| = |0 - 10^6| = 10^6 \rightarrow$ Very high error

So, choosing the first point as mean is a very bad method.