EEL6825: Pattern Recognition

Homework 2

Problem 4 in Chapter 3, on Page 141 of the textbook: Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", 2nd Edition, Wiley-Interscience, October 2000.

Problem 4 is given as below:

Let $\mathbf{x}$ be a d-dimensional binary (0 or 1) vector with a multivariate Bernoulli distribution

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

where $\boldsymbol{\theta} = (\theta_1, ..., \theta_d)^t$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Show that the maximum-likelihood estimate for $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k.$$

Ans. = Given that X is a d-dimensional binary vector.

let say we have n samples from the discrete distribution.

Samples are = $\{x_1, x_2, x_3 - - - - x_n\}$

The multivariate bernoulli distribution is given as =

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

Here $\theta = (\theta_1, \theta_2 \cdots \theta_d)^T$

θ_i = probablity that $x_i = 1$

for a perticular sequence among n-samples we can define the likelihood as =

$$P(x_1, x_2 \cdots x_n | \theta) = \prod_{k=1}^{n} \prod_{i=1}^{d} \theta_i^{x_{ki}} (1-\theta_i)^{1-x_{ki}}$$

Taking log both sides in above equation we obtain the log likelihood function =

$$l(\theta) = \sum_{k=1}^{n} \sum_{i=1}^{d} x_{ki} \ln \theta_i + (1-x_{ki}) \ln (1-\theta_i)$$

To find maximum - likelihood estimate for θ, we will have to put gradient of $l(\theta)$ to 0. which will give us θ at which function $l(\theta)$ is maximum.

Hence we have to put $\boxed{\nabla_\theta l(\theta) = 0}$

Also we will evaluate components for each value of ($i = 1, 2, \ldots d$)

lets evaluate gradient of function $l(\theta)$ with respect to θ.

$$\left[\nabla_{\theta} l(\theta)\right]_i = \nabla_{\theta_i} l(\theta) = 0$$

$$\Rightarrow \nabla_{\theta_i}\left[\sum_{k=1}^{n} x_{ki} \ln \theta_i + (1 - x_{ki}) \ln(1 - \theta_i)\right] = 0$$

(since we are valuating component by component)

$$\Rightarrow \frac{1}{\theta_i}\sum_{k=1}^{n} x_{ki} - \frac{1}{1-\theta_i}\sum_{k=1}^{n}(1 - x_{ki}) = 0$$

Hence we obtain, for any Particular i,

$$\frac{1}{\hat{\theta}_i}\sum_{k=1}^{n} x_{ki} = \frac{1}{1-\hat{\theta}_i}\sum_{k=1}^{n}(1 - x_{ki})$$

we can further solve this equation and rewrite it as =

$$(1-\hat{\theta}_i)\sum_{k=1}^{n} x_{ki} = \hat{\theta}_i\left(n - \sum_{k=1}^{n} x_{ki}\right)$$

$$\Rightarrow \sum_{k=1}^{n} x_{ki} - \cancel{\hat{\theta}_i \sum_{k=1}^{n} x_{ki}} = n\hat{\theta}_i - \cancel{\hat{\theta}_i \sum_{k=1}^{n} x_{ki}}$$

So we can write =

$$n\hat{\theta}_i = \sum_{k=1}^{n} x_{ki}$$

$$\hat{\theta}_i = \frac{1}{n}\sum_{k=1}^{n} x_{ki}$$

Since this is valid for all values of i. $i = (1, 2, \ldots d)$, we can write this equation in vector form as =

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

This $\hat{\theta}$ (maximum likelihood value of θ) is the sample mean value.