Quiz- II Exams; CS5011 –Machine Learning;
Time: 55 Mins.; Total Marks: 25; Mar. 2019.
Answer *only in the space provided below*. Answer all Questions.
Define all terms and Symbols used anywhere. Be brief and tidy.

ROLL: ________________: Name: ____________________________________

Q 1. (a) Give the equation of a multi-dimensional Gaussian Function; Define all the symbols used. **2 + 5 = 07**

$$p(X) = \frac{1}{\sqrt{\det(\Sigma)(2\pi)^d}} \exp[-\frac{(X-\mu)^T \Sigma^{-1}(X-\mu)}{2}]$$
$$= \frac{1}{\sqrt{\det(\Sigma)(2\pi)^d}} \exp[-\frac{1}{2}\sum_{ij}(x_i-\mu_i)s_{ij}(x_j-\mu_j)]$$

$$\boldsymbol{\Sigma - covariance\ matrix; d - dimension; X - data; \mu - mean\ vector}$$

where, s_{ij} is the i, j th component of Σ^{-1} (the inverse of covariance matrix $\mathbf{\Sigma}$).

(b) Find the Decision boundary for binary classification in 1-D case, under Bayes criteria, with identical class priors and Conditional PDF as Gaussians with unequal Variance (write only the main equations in derivation).

For the given conditions,

$$g_i(x) = -\frac{1}{2}\frac{(x-\mu_i)^2}{\sigma_i^2} - \frac{1}{2}\ln\sigma_i$$

For finding DB, $g_i(x) = g_j(x)$

$$\Rightarrow -\frac{1}{2\sigma_i^2}(x-\mu_i)^2 - \frac{1}{2}\ln\sigma_i^2 = -\frac{1}{2\sigma_j^2}(x-\mu_j)^2 - \frac{1}{2}\ln\sigma_j^2$$

$$\Rightarrow -\frac{1}{2}\cdot\left(\frac{1}{\sigma_i^2}(x^2+\mu_i^2-2x\mu_i) - \frac{1}{\sigma_j^2}(x^2+\mu_j^2-2x\mu_j) + \ln\sigma_i^2 - \ln\sigma_j^2\right) = 0$$

$$\Rightarrow -\frac{1}{2}\left(x^2\left(\frac{1}{\sigma_i^2}-\frac{1}{\sigma_j^2}\right) - x\left(\frac{2\mu_i}{\sigma_i^2}-\frac{2\mu_j}{\sigma_j^2}\right) + \left(\frac{\mu_i^2}{\sigma_i^2}+\ln\sigma_i^2-\frac{\mu_j^2}{\sigma_j^2}-\ln\sigma_j^2\right)\right) = 0$$

Representing as $(W_i - W_j)x^2 + (w_i - w_j)x + (w_{i0} - w_{j0}) = 0$

we get, $\boxed{W_i = \frac{-1}{2\sigma_i^2}\ ;\ w_i = \frac{\mu_i}{\sigma_i^2}\ \&\ w_{i0} = -\frac{1}{2}\frac{\mu_i^2}{\sigma_i^2} - \frac{1}{2}\ln\sigma_i^2}$

Q 2. In 2-D case for binary classification, find the expression of X_0, where the DB is defined as $(\omega_k^T - \omega_l^T)X + (\omega_{k0} - \omega_{l0}) = 0;$ **05**

and $(\omega_{k0} - \omega_{l0}) = (\omega_l - \omega_k)^T X_0$

where, LDF is $g_i(X) = \omega_i^T X + \omega_{i0}$; $\omega_i = \mu_i / \sigma^2$ and $\omega_{i0} = -\frac{\mu_i^T \mu_i}{2\sigma^2} + \ln P(w_i)$

$$(\omega_k^T - \omega_l^T) X + (\omega_{k0} - \omega_{l0}) = 0$$

$$\omega_i = \frac{\mu_i}{\sigma^2} \qquad \omega_{i0} = -\frac{\mu_i^T \mu_i}{2\sigma^2} + \ln P(\omega_i)$$

Sol:
$$\frac{1}{\sigma^2}(\mu_k - \mu_l)^T X - \frac{\mu_k^T \mu_k}{2\sigma^2} + \ln P(\omega_k) + \frac{\mu_l^T \mu_l}{2\sigma^2} - \ln P(\omega_l) = 0$$

Multiplying by σ^2 and re-arranging the terms:

$$(\mu_k - \mu_l)^T X - \frac{1}{2}(\mu_k^T \mu_k - \mu_l^T \mu_l) + \sigma^2 \ln\left(\frac{P(\omega_k)}{P(\omega_l)}\right) = 0$$

Taking $(\mu_k - \mu_l)^T$ common and rearranging:

$$(\mu_k - \mu_l)^T \left[X - \left\{ \frac{1}{2}(\mu_k + \mu_l) - \frac{\sigma^2 (\mu_k - \mu_l)}{\|\mu_k - \mu_l\|^2} \ln \frac{P(\omega_k)}{P(\omega_l)} \right\} \right] = 0$$

$\therefore\ W^t (X - X_0) = 0$

$W = \mu_k - \mu_l$

$$\boxed{X_0 = \frac{1}{2}(\mu_k + \mu_l) - \sigma^2 \frac{(\mu_k - \mu_l)}{\|\mu_k - \mu_l\|^2} \ln \frac{P(\omega_k)}{P(\omega_l)}}$$

Q 3. (a) In case of PCA, if, for data X: **2 + 3 = 05**

$$Y = W^T X = \Sigma V^T,$$

where $W\Sigma V^T$ is the singular v alue decomposit ion (SVD) of X.

and $$COV(X) = XX^T = W\Sigma\Sigma^T W^T = WDW^T$$

Find COV(Y):

$$COV(Y) = E[YY^T] = E[(W^T X)(W^T X)^T]$$
$$= E[(W^T X)(X^T W)] = W^T E[XX^T]W$$
$$= W^T COV(X)W = W^T (WDW^T)W = D$$

(b) Write the expression of LDA, by defining those of Within-class scatter and Between-class scatter matrices. Pl. Define all symbols and terms used.

$$W_{opt} = \arg\max \frac{|W^T S_B W|}{|W^T S_W W|}$$

Where within class scatter: $$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

Between class scatter: $$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T$$

$\mu_i - mean\ of\ class\ i, \mu - mixture\ mean, N_i - number\ of\ points\ in\ class\ i;$

$W - direction\ along\ which\ X\ is\ projected\ in\ order\ to\ get\ maximum$

$interclass\ variance\ and\ minimum\ intra\ class\ variance.$

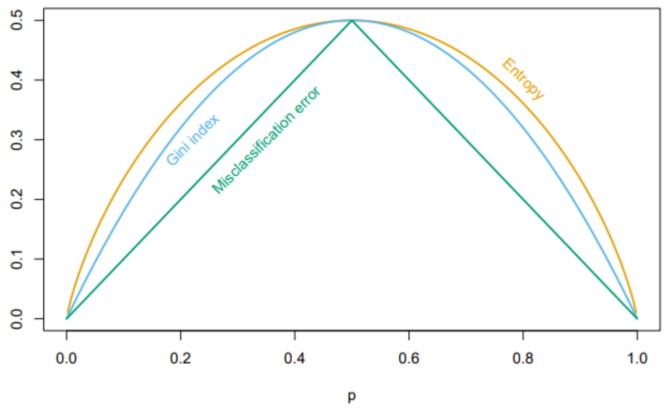
Q 4. (a) Define the term GINI index; again define all variables/terms used
2 + 3 = 05

GINI INDEX: $$Q_\tau(T) = \sum_{k=1}^{K} p_{\tau k}\left(1 - p_{\tau k}\right)$$

we define $p_{\tau k}$ to be the proportion of data points in region R_τ assigned to class k, where $k = 1, \ldots, K$ with K being the total number of classes. The leaf nodes are indexed by $\tau = 1, \ldots, |T|$, with leaf node τ representing a region R_τ.

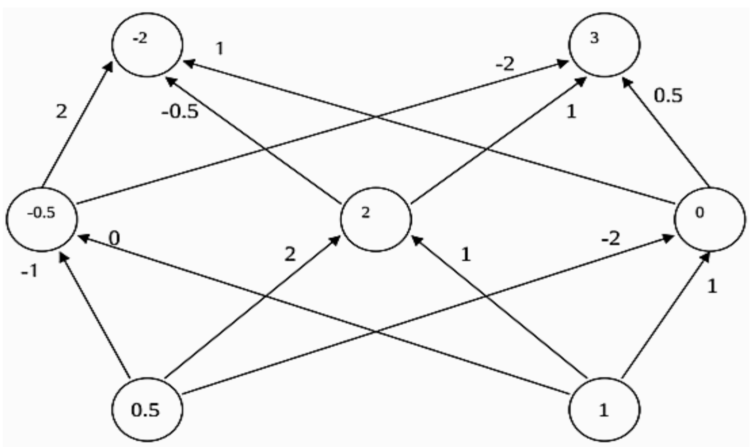
Gini index: $$\sum_{k \neq k'} \hat{p}_{mk}\hat{p}_{mk'} = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}).$$

(b) Plot the GINI index vs probability of class samples (for two-class problem). If the maximum of this plot larger than ENTROPY computed for the same scenario?



No, it is not.

Q 5. The following is a network of linear neurons. The numbers at connections indicate the value of the corresponding weight, and the numbers in the circles indicate the output of a neuron, for a given input (0.5, 1). 03



What is the output of the network if the input changes to (1, 2)? Is there an easy way to compute this specific case- explain with one sentence?

Output for (1, 2) = (-4, 6). (Because of the **linearity**).

As the input has been doubled, keeping their ratio intact, the corresponding output will also be doubled.