1.1 Minimizing Error:

Here a complete orthonormal set of D-dimensional basis vectors $\{u_i\}$ where $i = 1, \dots, D$ that satisfy

$$u_i^T u_j = \delta_{ij}.$$

Each data point can be represented exactly by a linear combination of the basis vectors

$$x_n = \sum_{i=1}^{D} \alpha_{ni} u_i$$

the coefficients α_{ni} will be different for different data points. The original D components $\{x_{n1}, \dots, x_{nD}\}$ are replaced by an equivalent set $\{\alpha_{n1}, \dots, \alpha_{nD}\}$. Taking the inner product with u_j and making use of the orthonormality property, we obtain $\alpha_{nj} = x_n^T u_j$ and without loss of generality we can write

$$x_n = \sum_{i=1}^{D} (x_n^T u_i) u_i.$$

This data point using a representation involving a restricted number $M < D$ of variables corresponding to a projection onto a lower-dimensional subspace. The M-dimensional linear subspace can be represented without loss of generality, by the first M of the basis vectors and each data point x_n by

$$\bar{x}_n = \sum_{i=1}^{M} z_{ni} u_i + \sum_{i=M+1}^{D} b_i u_i$$

where the $\{z_{ni}\}$ depend on the particular data point, whereas the $\{b_i\}$ are constants that are the same for all the data points. So, we can choose the $\{u_i\}$, $\{z_{ni}\}$ and $\{b_i\}$ so as to minimize the distortion introduced by the reduction in dimensionality.

The goal is to minimize, $J = \frac{1}{N} \sum_{n=1}^{N} \| x_n - \bar{x}_n \|^2$

Considering first of all the minimization with respect to the quantities $\{z_{ni}\}$. Substituting for $\bar{x}_n$, setting the derivative with respect to z_{nj} to zero we get,

$$z_{nj} = x_n^T u_j$$

where $j = 1, \ldots, M$. Similarly, setting the derivative of J with respect to b_i to zero, $b_j = \bar{x}^T u_j$

where, $j = M+1, \ldots, D$. If we substitute for z_{ni} and b_i, we get

$$x_n - \bar{x}_n = \sum_{i=M+1}^{D} \{(x_n - \bar{x})^T u_i\} u_i$$

So,

$$J = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=M+1}^{D} (x_n^T u_i - \bar{x}^T u_i)^2$$

$$= \sum_{i=M+1}^{D} u_i^T S u_i$$

To minimize J we choose u_2 direction and subject to the normalization constraint $u_2^T u_2 = 1$. Using a Lagrange multiplier λ_2 to enforce the constraint, we consider the minimization of,

$$J = u_2^T S u_2 + \lambda_2 (1 - u_2^T u_2)$$

The general solution to the minimization of J for arbitrary D and arbitrary M < D is obtained by choosing the $\{u_i\}$ to be eigenvectors and covariance matrix given by

$$S u_i = \lambda_i u_i$$

where, $i = 1, \ldots, D$ and eigenvectors $\{u_i\}$ are chosen to be orthonormal.

$$J = \sum_{i=M+1}^{D} \lambda_i$$

The minimum value of J by selecting these eigenvectors to be those having the D-M smallest eigenvalue and the eigenvectors defining the principle subspace corresponding to the M largest eigenvalues.

2.1

Likelihood expression of a class k with prior probability π_k using Bayes' rule:

$$\cancel{P_k(x) = \frac{p(x|k)\,p(k)}{p(x)}} \qquad P(k|x) = \frac{P_k(x)\,\pi_k}{P(x)}$$

Theorem gives us:

$$Pr(G=k \mid X=x) = \frac{f_k(x)\,\pi_k}{\sum_{l=1}^{k} f_l(x)\,\pi_l}$$

modelling each class density as multivariate Gaussian.

$$P_k(x) = \frac{1}{(2\pi)^{p/2}\,|\Sigma_k|^{1/2}}\, e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)}$$

Linear discriminant analysis arises in the special case when the classes have a common covariance matrix $\Sigma_k = \Sigma \;\forall k$. So the log-ratio;

$$\log \frac{Pr(G=k \mid X=x)}{Pr(G=l \mid X=x)} = \log \frac{P_k(x)}{P_l(x)} + \log \frac{\pi_k}{\pi_l}$$

$$= \log \frac{\pi_k}{\pi_l} - \frac{1}{2}(\mu_k + \mu_l)^T \Sigma^{-1} (\mu_k - \mu_l) + x^T \Sigma^{-1} (\mu_k - \mu_l) \quad \text{— (1)}$$

From eqn (1), LD functions.

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2}\mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

The discriminant functions:-

$$\delta_k(x) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k) + \log \pi_k$$