Discriminant Analysis

Discriminant Function:We want to separate out distinct population in an optimal way. These are multidimensional and we want a way to do so!

Classification: Will allocate as per the discriminant function into appropriate population.

Basically train and test the data to classify properly without overfitting!

Underlying Data Structure:

Let $\{x_1 \ldots x_n\}$ = multidimensional feature vectors

Suppose that there are J class memberships, such that C = {1… J}

Where each of the feature vectors will have a class membership assigned to it. Therefore there should be a rule that assigns a feature vector to a particular membership, sort of like a prediction.

Let's start with a 2 class classification problem $\pi_1 \; and \; \pi_2$

Let x_1 ~ π_1 and π_2 ~ π_2; our goal is to find a function g(x1) which looks completely different from g(x2) as they come from entirely different populations. Then ‘g’ becomes our discriminant function.

Then, when we encounter a new ‘x’, it can be classified by ‘g’.

Fisher LDA

Assumes:

$x|\pi_1$, then we characterize it with the mean μ_1 and covariance matrix Σ

$x|\pi_2$, then we characterize it with the mean μ_2 and covariance matrix Σ

But as we know that the x we have here is multivariate and therefore need to convert this into a univariate case!

How do we do this? - Take a linear combination by replacing x with l'x.

Therefore we can say:

$l'x|\pi_1$, then we characterize it with the mean $l'\mu_1$ and covariance matrix $l'\Sigma l$

$l'x|\pi_2$, then we characterize it with the mean $l'\mu_2$ and covariance matrix $l'\Sigma l$

This 'l' is the parameter that we need to tune in order to make the distance between the 2 classes as large as possible since the covariance of both π_1 *and* π_2 is the same.

The statistical distance between the 2 populations π_1 *and* π_2 is given by:

$$[\ (l'\mu_1 - l'\mu_2)^2 / \ l'\Sigma l\ ] = [\ l'(\mu_1 - \mu_2)^2 / \ l'\Sigma l\ ]$$

Therefore we are trying to max_l' ([$l'(\mu_1 - \mu_2)^2 / \ l'\Sigma l$])

On further simplifying, and making use of the cauchy schwarz inequality,

$[\, l'(\mu_1 - \mu_2)^2 / \; l'\Sigma l \,] <= (\mu_1 - \mu_2)' \Sigma' (\mu_1 - \mu_2)$ [This is the Mahalanobis Distance]

We can further simplify this to get the 'l' we are looking for:

$$l = \Sigma^{-1} (\mu_1 - \mu_2)'$$

This is the Linear Fisher Discriminant Function:

$$l' X = \Sigma^{-1} (\mu_1 - \mu_2)' X$$

To find out how the points are being classified, we need to look at the expectation of the function.

Therefore, $E[\, \Sigma^{-1} (\mu_1 - \mu_2)' X \mid \pi_i \,] = \Sigma^{-1} (\mu_1 - \mu_2)' X * \mu_i = m_i$

Note that if we do the following for the 2 class classification case,

$m_1 - m_2 = (\mu_1 - \mu_2)' \Sigma' (\mu_1 - \mu_2) >= 0$, this is always going to be true since Σ' is positive definite and of = 0, belong to the same class! Ta da…

Therefore m1 >= m2 and therefore a point can be assigned to either of the classes of closer to m1 or m2. This can be understood from the Expectation that we had taken before.

Note: We need to take the sample means and variances since they are almost always not known for the data.

COST OF MISCLASSIFICATION:

C(i|j) when a point from π_j is misclassified in π_i.