Machine Learning

Linear Discriminant Analysis (LDA)

Digital Vidya

Linear Discriminant Analysis

- The purpose of Discriminant Analysis is to perform Dimensionality reduction

- In general, we assign an object to one of a number of predetermined groups based on observations made on the object.

- For example, we want to know whether a soap product is good or bad based on several measurements on the product such as weight, volume, people's preferential score, smell, color contrast etc. The object here is soap. The class category or the group ("good" and "bad") is what we are looking for (it is also called dependent variable). Each measurement on the product is called features that describe the object (it is also called independent variable)

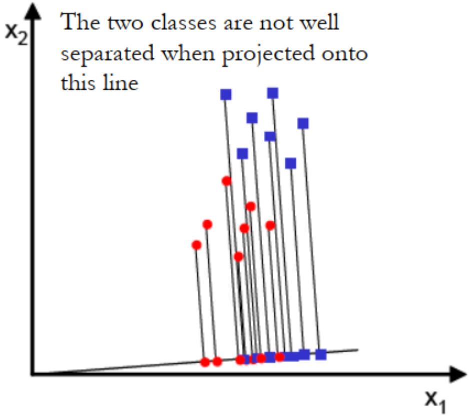
Linear Discriminant Analysis

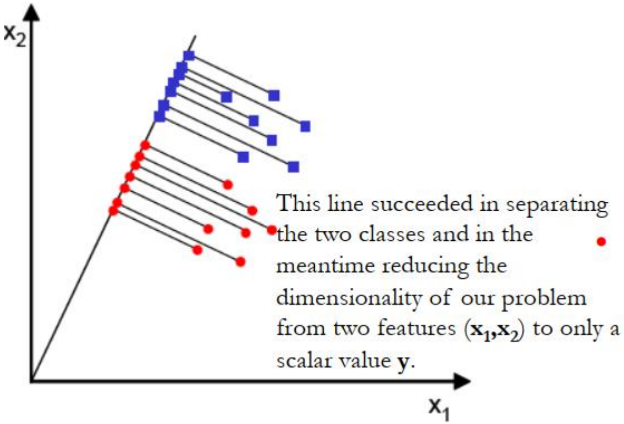
- In discriminant analysis, the dependent variable (Y) is the group and the independent variables (X) are the object features that might describe the group. The dependent variable is always category (nominal scale) variable while the independent variables can be any measurement scale (i.e. nominal, ordinal, interval or ratio)

- If the groups are linearly separable, we can use linear discriminant model (LDA)

- Linearly separable suggests that the groups can be separated by a linear combination of features that describe the objects. If only two features, the separators between objects group will become lines. If the features are three, the separator is a plane and the number of features (i.e. independent variables) is more than 3, the separators become a hyper-plane.

LDA … Two Classes





- Assume we have *m*-dimensional samples {**x1, x2,…, xN**}, *N1* of which belong to **ω1** and *N2* belong to **ω2**.
- We seek to obtain a scalar **y** by projecting the samples **x** onto a line (C-1 space, C = 2)

$$y = w^T x \quad where \quad x = \begin{bmatrix} x_1 \\ . \\ . \\ x_m \end{bmatrix} \quad and \quad w = \begin{bmatrix} w_1 \\ . \\ . \\ w_m \end{bmatrix}$$

- where **w** is the projection vectors used to project **x** to **y**
  Of all the possible lines we would like to select the one that maximizes the separability of the scalars.

Digital Vidya

LDA ... Two Classes

- In order to find a good projection vector, we need to define a measure of separation between the projections
- The mean vector of each class in x and y feature space is:

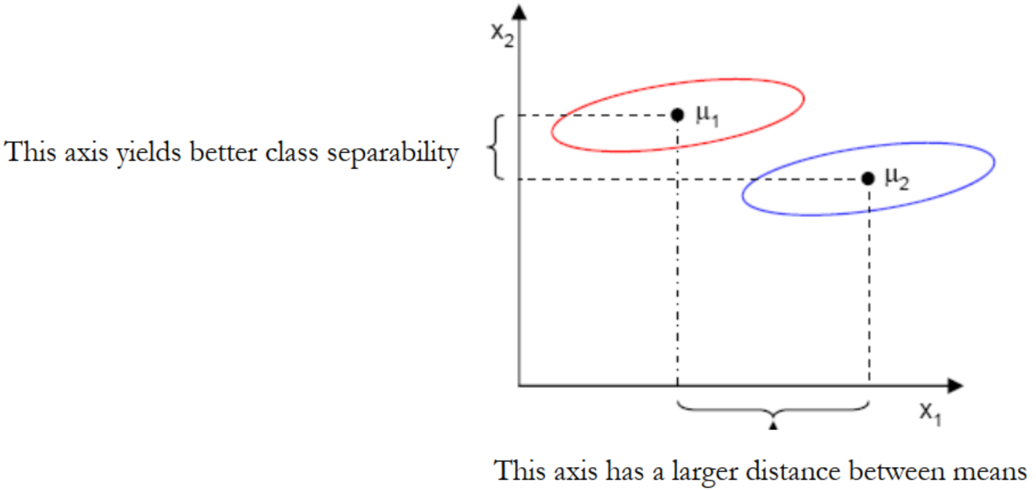
$$\mu_i = \frac{1}{N_i}\sum_{x\in\omega_i} x \quad and \quad \tilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y = \frac{1}{N_i}\sum_{x\in\omega_i} w^T x$$
$$= w^T \frac{1}{N_i}\sum_{x\in\omega_i} x = w^T \mu_i$$

- – i.e. projecting x to y will lead to projecting the mean of x to the mean of y
- We could then choose the distance between the projected means as our objective function

$$J(w) = \left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|w^T\mu_1 - w^T\mu_2\right| = \left|w^T(\mu_1 - \mu_2)\right|$$

LDA ... Two Classes

- However, the distance between the projected means is not a very good measure since it does not take into account the standard deviation within the classes



LDA ... Two Classes

- The solution proposed by Fisher is to maximize a function that represents the difference between the means, normalized by a measure of the within-class variability, or the so-called *scatter.*
- For each class we define the **scatter**, an equivalent of the variance, as
  (sum of square differences between the projected samples and their class mean)

$$\tilde{s}_i^2 = \sum_{y \in \omega_i} (y - \tilde{\mu}_i)^2$$

- $\tilde{s}_i^2$ measures the variability within class **ωi** after projecting it on the y-space
- Thus measures the variability within the two classes at hand after projection, hence it is called *within-class scatter* of the projected samples

LDA ... Two Classes

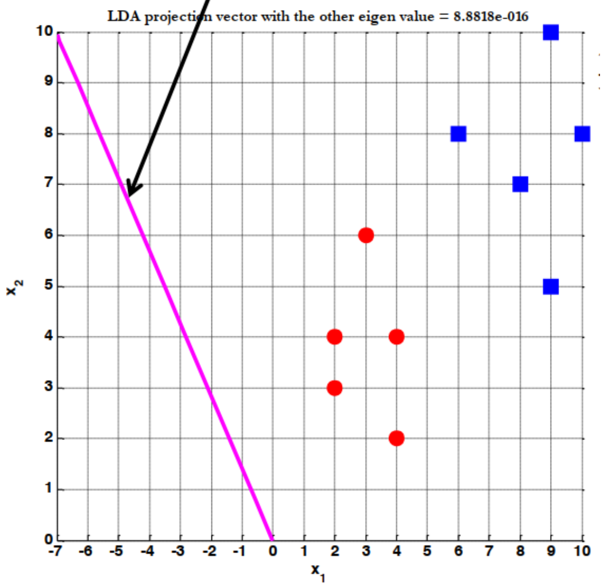
- We can finally express the Fisher criterion in terms of SW and SB as:

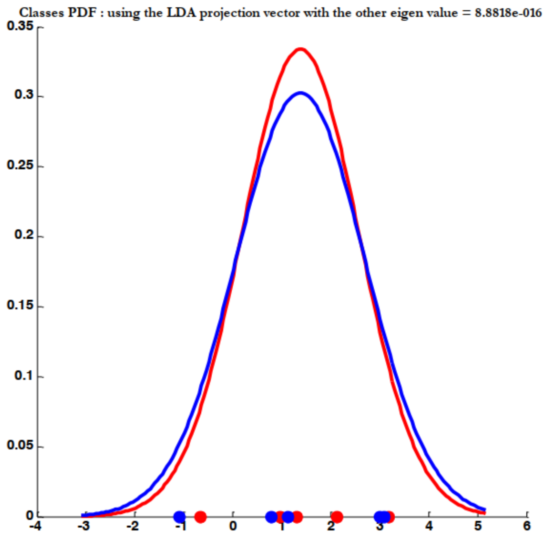
$$\boxed{J(w) = \frac{\left|\tilde{\mu}_1 - \tilde{\mu}_2\right|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{w^T S_B w}{w^T S_W w}}$$

- Hence J(w) is a measure of the difference between class means (encoded in the between-class scatter matrix) normalized by a measure of the within-class scatter matrix

LDA Projection

The projection vector corresponding to the **smallest** eigen value

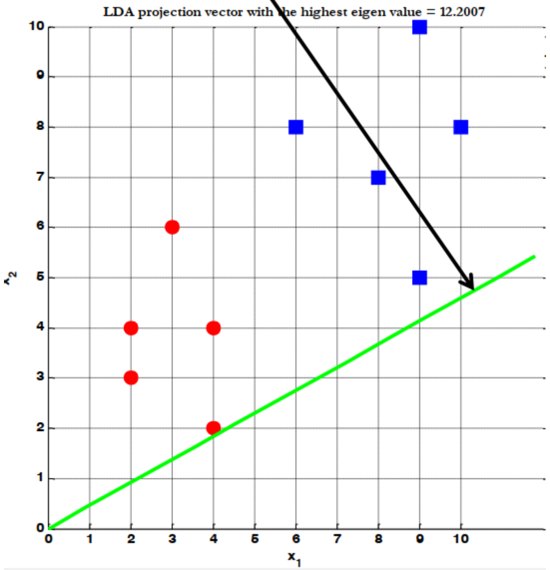


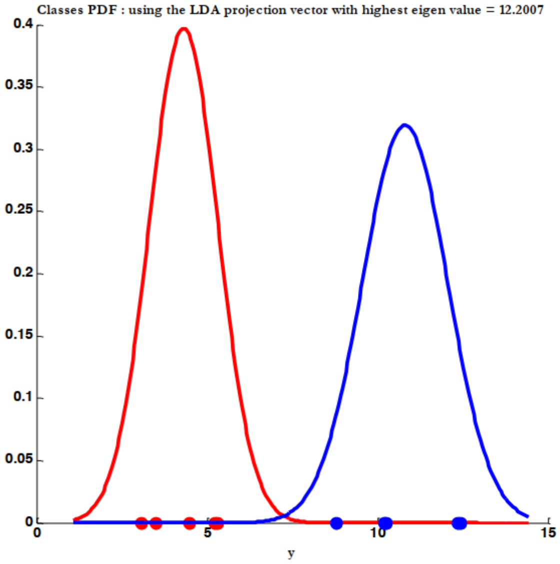


Using this vector leads to **bad separability** between the two classes

LDA Projection

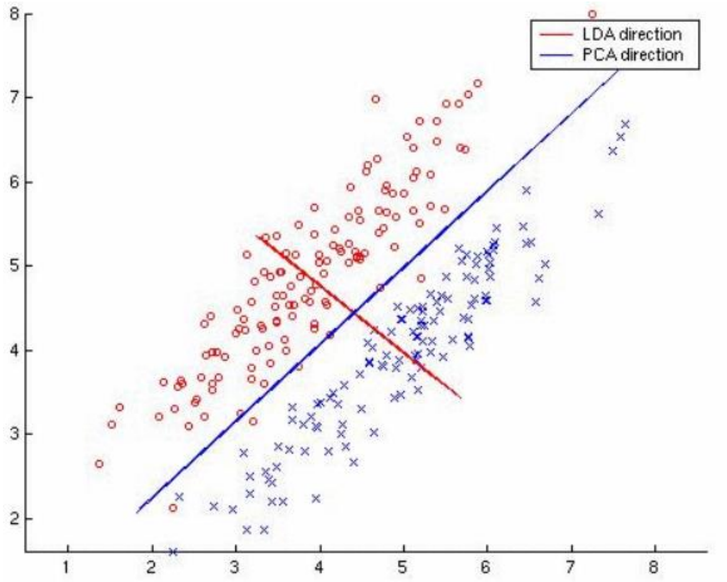
The projection vector corresponding to the **highest** eigen value



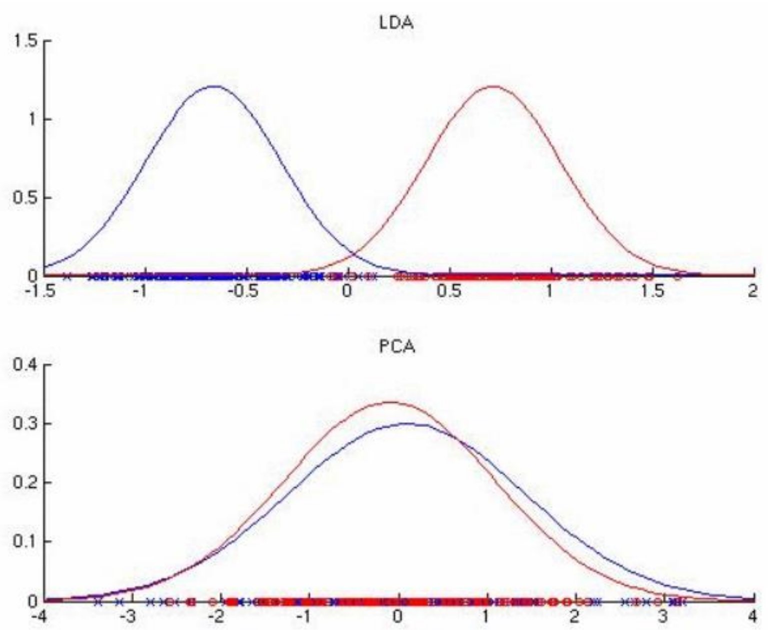


Using this vector leads to **good separability** between the two classes

PCA vs LDA



LDA direction
PCA direction
8
7
6
5
4
3
2
1
2
3
4
5
6
7
8



LDA
1.5
1
0.5
0
-1.5
-1
-0.5
0
0.5
1
1.5
2
PCA
0.4
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3
4

Thank You

Digital Vidya