

Minimum Distance Classifiers

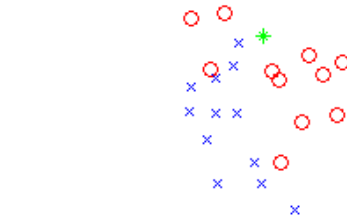
Notations and problem setting

(For basic definitions, go to [this page](#).)

We have a dataset D containing n sample vectors of dimension d . Some of these sample vectors are labelled, some are not. We now want to classify (label) the unlabeled sample vectors, given the labelled ones.

The method will be to compute the mean of each class and measure the distance to the unlabelled vector according to some metric. The class with a mean closest to the vector (with smallest distance) will be the label for that vector.

For definitions of the different metrics, go to [this page](#).



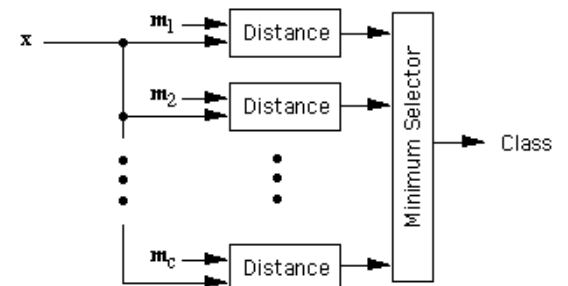
2D example: given sample vectors with red (circles) and blue (x) labels, what is the label of the green one (cross)?

The minimum distance to the classes' means

The first step is to compute, for each kind of label L_i found in the dataset, the mean of all sample vectors that were labelled with that particular label. Then, the distance between each unlabelled sample vector x we want to classify, and the means of all labelled sample vectors is minimized to give the class. If we would use the Euclidean distance metric for instance, we would find the class L by:

$$L = \underset{L_i}{\operatorname{argmin}} \sqrt{\sum_{j=1}^d (x_j - \mu_j(L_i))^2}$$

The idea behind this is that the mean should be a representative value for the class, defining usually the centre of all the sample vectors that were labelled as that class in input space.



schematic overview of the minimum distance procedure, where m stands for the mean of a certain class

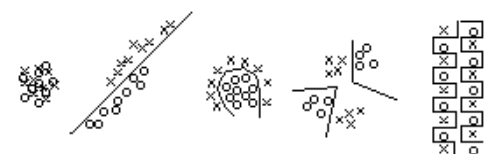
There are limitations in using the Euclidean distance metric, however, since it is linear in nature:

1. The features may be inadequate to distinguish the different classes
2. The features may be highly correlated
3. The decision boundary may have to be curved
4. There may be distinct subclasses in the data
5. The feature space may simply be too complex

Using the Mahalanobis distance metric helps in most of these cases.

The maximum correlation classifier

The Euclidean formula can be rewritten as:



illustrations of the limitations 1 - 5 of linear classifiers

$$\begin{aligned}\sum_{j=1}^d (x_j - \mu_j(L_i))^2 &= (\vec{x} - \vec{\mu}(L_i))' \cdot (\vec{x} - \vec{\mu}(L_i)) \\ &= -2 \left[\vec{\mu}(L_i)' \vec{x} - \frac{1}{2} \vec{\mu}(L_i)' \cdot \vec{\mu}(L_i) \right] + \vec{x}' \vec{x}\end{aligned}$$

so minimizing the first term will be equivalent to maximizing the term between the square brackets, this will lead to a new function:

A linear discriminant function is defined as:

$$\begin{aligned}g(\vec{x}, L_i) &= \vec{\mu}(L_i)' \vec{x} - \frac{1}{2} \vec{\mu}(L_i)' \cdot \vec{\mu}(L_i) \\ &= \sum_{j=1}^d (\mu_j(L_i) x_j) - \frac{1}{2} \sum_{j=1}^d (\mu_j(L_i))^2\end{aligned}$$

we thus take the maximum g value instead of taking the minimum Euclidean distance. This is also called maximizing the correlation (as opposed to the minimizing the distance).

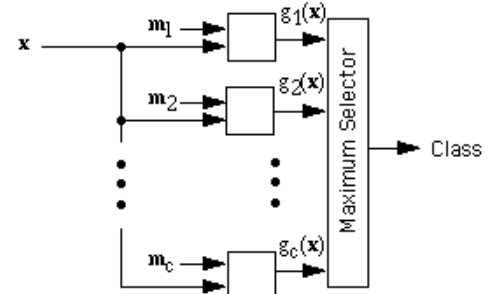
An invariant linear discriminant function can be obtained from the Mahalanobis metric if the assumption is made that the covariance matrix is the same for all classes:

$$\begin{aligned}&(\vec{x} - \vec{\mu}(L_i))' \cdot \text{Cov}(L)^{-1} \cdot (\vec{x} - \vec{\mu}(L_i)) \\ &= \vec{x}' \cdot \text{Cov}(L)^{-1} \vec{x} - \vec{\mu}(L_i)' \cdot \text{Cov}(L)^{-1} \vec{x} - \vec{x}' \cdot \text{Cov}(L)^{-1} \cdot \vec{\mu}(L_i) + \vec{\mu}(L_i)' \cdot \text{Cov}(L)^{-1} \cdot \vec{\mu}(L_i) \\ &= -2 \left[\left(\text{Cov}(L)^{-1} \cdot \vec{\mu}(L_i) \right)' \vec{x} - \frac{1}{2} \vec{\mu}(L_i)' \cdot \text{Cov}(L)^{-1} \cdot \vec{\mu}(L_i) \right] + \vec{x}' \cdot \text{Cov}(L)^{-1} \cdot \vec{x}\end{aligned}$$

The invariant linear discriminant function is then:

$$g_{inv}(\vec{x}, L_i) = \left(\text{Cov}(L_i)^{-1} \cdot \vec{\mu}(L_i) \right)' \vec{x} - \frac{1}{2} \vec{\mu}(L_i)' \cdot \text{Cov}(L_i)^{-1} \cdot \vec{\mu}(L_i)$$

Boundaries are now linear, but they are invariant to linear transformations, plus the calculations are simpler as well.



schematic overview of the maximal correlation procedure, where m stands for the mean of a certain class g can be a linear or invariant linear discriminant function

Compiled by [Kristof Van Laerhoven](#).

Most came straight from the following web pages:

[Klassifikation Genkendelse og Tracking](#)

[PR home](#)