Master of Information and Data Science
DATASCI 281: Computer Vision
Spring 2022

Professor Hany Farid
University of California, Berkeley

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a standard approach to multi-class classification. LDA projects data onto a linear subspace so that the within-class scatter (specifically, the within-class variance) is minimized and the across-class scatter is maximized. Shown in Figure 1 are two classes of data in a 2-D space (red and blue solid points), and their projection onto an axis that is optimal for classification (red and blue open circles).

Note that in this 1-D space, the data are perfectly separated and can be classified with a simple threshold. Novel data are projected onto the same axis and classified by comparing against a threshold.

For simplicity a two-class LDA is described – the extension to multiple classes is straight-forward. Denote column vectors $\vec{x}_i$, $i = 1, ..., N_x$ and $\vec{y}_j$, $j = 1, ..., N_y$ as training data from each of two classes. The within-class means are defined as:

$$\vec{\mu}_x = \frac{1}{N_x} \sum_{i=1}^{N_x} \vec{x}_i, \quad \text{and} \quad \vec{\mu}_y = \frac{1}{N_y} \sum_{j=1}^{N_y} \vec{y}_j. \tag{1}$$

The between-class mean is defined as:

$$\vec{\mu} \;\; = \;\; \frac{1}{N_x + N_y} \left(\sum_{i=1}^{N_x} \vec{x}_i + \sum_{j=1}^{N_y} \vec{y}_j \right). \tag{2}$$

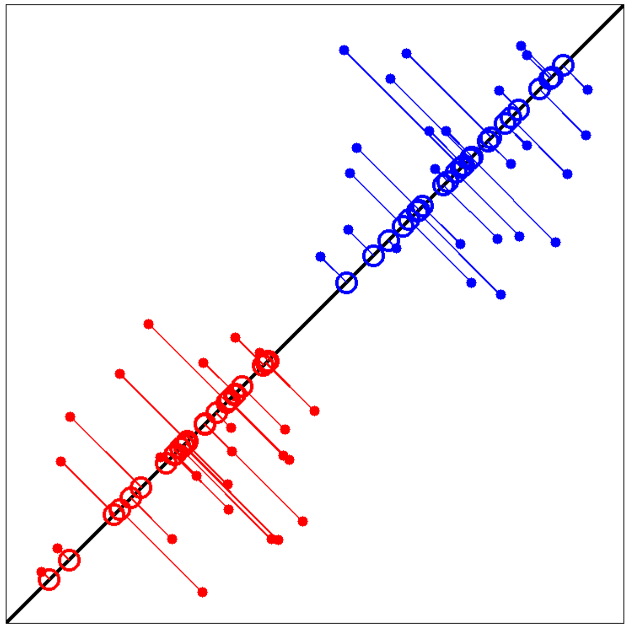

Figure 1: Two class LDA.

The within-class scatter matrix is defined as:

$$S_w \quad = \quad M_x M_x^T + M_y M_y^T, \tag{3}$$

where, the i^{th} column of matrix M_x contains the zero-meaned i^{th} exemplar given by $\vec{x}_i - \vec{\mu}_x$. Similarly, the j^{th} column of matrix M_y contains $\vec{y}_j - \vec{\mu}_y$. The between-class scatter matrix is defined as:

$$S_b \quad = \quad N_x(\vec{\mu}_x - \vec{\mu})(\vec{\mu}_x - \vec{\mu})^T + N_y(\vec{\mu}_y - \vec{\mu})(\vec{\mu}_y - \vec{\mu})^T. \tag{4}$$

Let $\vec{e}$ be the maximal generalized eigenvalue-eigenvector of S_b and S_w (i.e., $S_b\vec{e} = \lambda S_w \vec{e}$). The training data $\vec{x}_i$ and $\vec{y}_j$ are projected onto the one-dimensional linear subspace defined by $\vec{e}$ (i.e., $\vec{x}_i^T\vec{e}$ and $\vec{y}_j^T\vec{e}$). This projection simultaneously minimizes the within-class scatter while maximizing the between-class scatter.

Once the LDA projection axis is determined from the training set, a novel exemplar, $\vec{z}$, from the testing set is classified by first projecting onto the same subspace, $\vec{z}^T\vec{e}$. In the simplest case, the class to which this exemplar belongs is determined via a simple threshold.

In the case of a two-class LDA, we are guaranteed to be able to project onto a one-dimensional subspace (i.e., there will be at most one non-zero eigenvalue). In the case of a N-class LDA, the projection may be onto as high as a $N-1$-dimensional subspace.

By formulating the optimization in terms of maximizing projected variance, it is being implicitly assumed that the original data is Gaussian distributed. Significant deviations of data from this assumption can result in poor classification results.