FISHER'S LINEAR DISCRIMINANT ANALYSIS

Fisher's Linear Discriminant Analysis (FDA) is a dimensionality reduction technique to ease classification.

1 Two Class Case

Consider the case of N points in d dimensions, with each point belonging to one of the two classes C_1 and C_2. The idea is to find the optimal direction to project the vector of these points to. Such a projection can be represented as,

$$y = \boldsymbol{w}^T \boldsymbol{x}$$

where $\boldsymbol{w}$ is a d dimensional vector defining the direction of projection, $\boldsymbol{x}$ is the vector being projected and y is a scalar representing the magnitude of projection.

The projection should serve two purposes as discussed in the following subsections:

1.1 Maximizing Between-Class Scatter

The class means should be projected as far apart as possible. Let $\boldsymbol{m_1}$ and $\boldsymbol{m_2}$ denote the mean of class C_1 and C_2 respectively.

$$\boldsymbol{m_1} = \frac{1}{N} \sum_{\boldsymbol{x_i} \in C_1} \boldsymbol{x_i}$$
$$\boldsymbol{m_2} = \frac{1}{N} \sum_{\boldsymbol{x_i} \in C_2} \boldsymbol{x_i}$$

If $\boldsymbol{m_1}$ is projected to m_1 and $\boldsymbol{m_2}$ is projected to m_2, then the between-scatter is defined by,

$$
\begin{aligned}
(m_1 - m_2)^2 &= (\boldsymbol{w}^T \boldsymbol{m_1} - \boldsymbol{w}^T \boldsymbol{m_2})^2 \\
&= (\boldsymbol{w}^T (\boldsymbol{m_1} - \boldsymbol{m_2}))^2 \\
&= \boldsymbol{w}^T (\boldsymbol{m_1} - \boldsymbol{m_2}) \boldsymbol{w}^T (\boldsymbol{m_1} - \boldsymbol{m_2}) \\
&= \boldsymbol{w}^T (\boldsymbol{m_1} - \boldsymbol{m_2}) (\boldsymbol{m_1} - \boldsymbol{m_2})^T \boldsymbol{w} \\
&= \boldsymbol{w}^T \boldsymbol{S_B} \boldsymbol{w}
\end{aligned}
$$

where $\boldsymbol{S_B}$ represents the between-class scatter matrix.

1.2 Minimizing Within-Class Scatter

The projections of each class should be as condensed as possible. The within-class scatter of the transformed data belonging to class C_k is denoted by,

$$
\begin{aligned}
s_k^2 &= \sum_{i \in C_k} (y_i - m_k)^2 \\
&= \sum_{i \in C_k} (\boldsymbol{w}^T \boldsymbol{x_i} - \boldsymbol{w}^T \boldsymbol{m_i})^2 \\
&= \sum_{i \in C_k} (\boldsymbol{w}^T (\boldsymbol{x_i} - \boldsymbol{m_i}))^2 \\
&= \sum_{i \in C_k} \boldsymbol{w}^T (\boldsymbol{x_i} - \boldsymbol{m_i})(\boldsymbol{x_i} - \boldsymbol{m_i})^T \boldsymbol{w} \\
&= \boldsymbol{w}^T \boldsymbol{S_k} \boldsymbol{w}
\end{aligned}
$$

where,

$$
\boldsymbol{S_k} = \sum_{i \in C_k} (\boldsymbol{x_i} - \boldsymbol{m_i})(\boldsymbol{x_i} - \boldsymbol{m_i})^T
$$

In the case of two classes, total within class scatter is denoted by,

$$
\begin{aligned}
s_1^2 + s_2^2 &= \boldsymbol{w}^T \boldsymbol{S_1} \boldsymbol{w} + \boldsymbol{w}^T \boldsymbol{S_1} \boldsymbol{w} \\
&= \boldsymbol{w}^T (\boldsymbol{S_1} + \boldsymbol{S_2}) \boldsymbol{w} \\
&= \boldsymbol{w}^T \boldsymbol{S_W} \boldsymbol{w}
\end{aligned}
$$

where $\boldsymbol{S_W}$ represents the within-class scatter matrix.

1.3 Combining Minimization and Maximization

A reasonable way to simultaneously maximize the between-class scatter and minimize the within-class scatter is to maximize their fraction defined as follows,

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \boldsymbol{S_B} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S_W} \boldsymbol{w}}$$

Note that $J(\boldsymbol{w})$ is invariant under rescalings of the form $\boldsymbol{w} \Rightarrow \alpha \boldsymbol{w}$. This sets up the reformulation of this problem as per the following,

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{w}^T \boldsymbol{S_B} \boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^T \boldsymbol{S_W} \boldsymbol{w} = 1 \end{aligned}$$

Using the concept of Langrangian,

$$L(\boldsymbol{w}, \lambda) = \boldsymbol{w}^T \boldsymbol{S_B} \boldsymbol{w} - \lambda(\boldsymbol{w}^T \boldsymbol{S_W} \boldsymbol{w} - 1)$$

Differentiating w.r.t. $\boldsymbol{w}$,

$$\frac{\partial L(\boldsymbol{w}, \lambda)}{\partial \boldsymbol{w}} = 2\boldsymbol{S_B} \boldsymbol{w} - 2\lambda \boldsymbol{S_W} \boldsymbol{w}$$

Equating the diffenrential to zero, the solution follows,

$$\boldsymbol{S_B} \boldsymbol{w} = \lambda \boldsymbol{S_W} \boldsymbol{w}$$

This is the generalized eigenvalue problem that can be solved easily. Note that since $\boldsymbol{S_B}$ is a product of two vectors and thus of rank one, the above equation will only yield one eigenvalue, eigenvector pair. The eigenvector is the sought projection direction.

2 Multi-Class Case