Fisher Discriminant

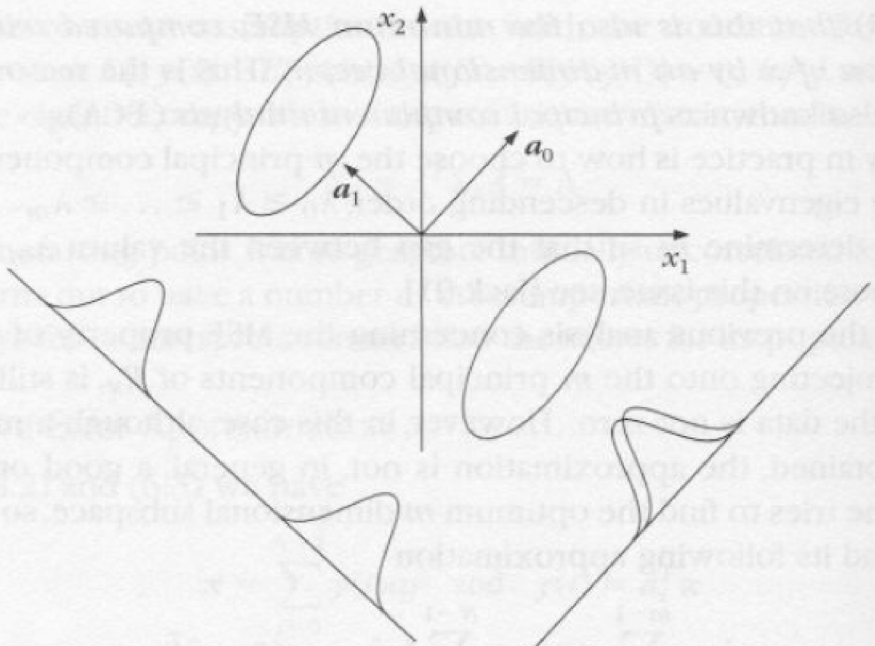

FIGURE 6.1

The KL transform is not always best for pattern recognition. In this example, projection on the eigenvector with the larger eigenvalue makes the two classes coincide. On the other hand projection on the other eigenvector keeps the classes separated.

Fisher's Discriminant

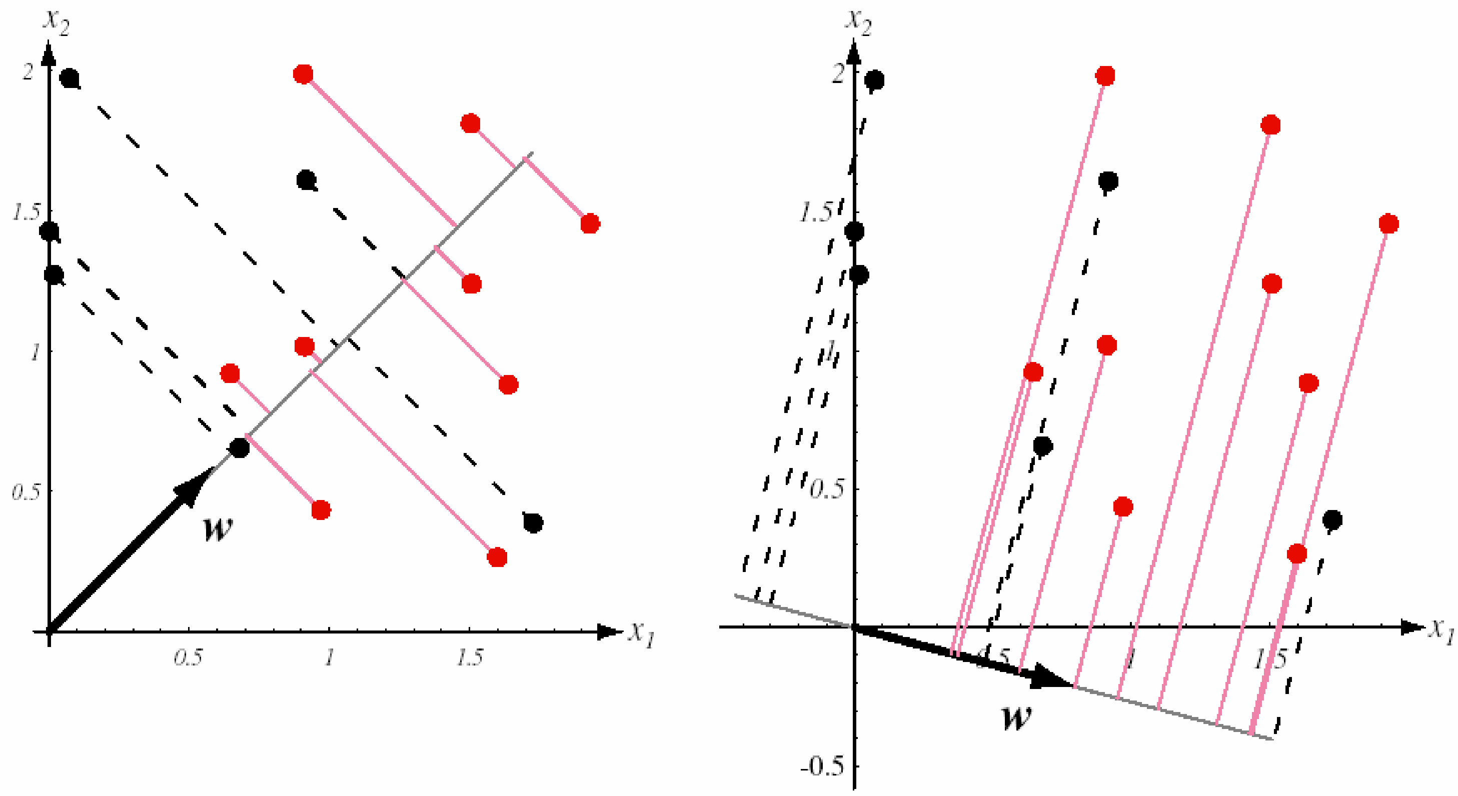

FIGURE 3.5. Projection of the same set of samples onto two different lines in the directions marked **w**. The figure on the right shows greater separation between the red and black projected points. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Fisher's Discriminant for two category case

Sample mean for the class ω_i

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x} \qquad i = 1,2$$

Individual sample mean for class ω_i after projection

$$\tilde{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{w}^T \mathbf{x} = \sum_{y \in D_i} y = \mathbf{w}^T \mathbf{m}_i$$

Sample scatter / covariance matrix for the class ω_i

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

Fisher's Discriminant for two category case

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

Fisher's discriminant criterion function needs to be maximized in this algorithm

$$J(\mathbf{w}) = \frac{\left|\tilde{m}_1 - \tilde{m}_2\right|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

Fisher's Discriminant for two category case

$$\tilde{s}_1^2 + \tilde{s}_2^2 = \mathbf{w}^T \mathbf{S}_1 \mathbf{w} + \mathbf{w}^T \mathbf{S}_2 \mathbf{w} = \mathbf{w}^T \mathbf{S}_{\mathbf{w}} \mathbf{w}$$

Similarly

$$\begin{aligned}(\tilde{m}_1 - \tilde{m}_2)^2 &= (\mathbf{w}^T \mathbf{m}_1 - \mathbf{w}^T \mathbf{m}_2)^2 \\ &= \mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{S}_B \mathbf{w}\end{aligned}$$

where

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$$ **Rank = 1**

$\mathbf{S_w}$ is called within class scatter matrix and $\mathbf{S}_\mathrm{B}$ is called between class scatter matrix.

Fisher's Discriminant for two category case

Let us define

$$\mathbf{S}_i = \sum_{\mathbf{x}\in D_i} (\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T \quad \text{and} \quad \mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2$$

Since $y = \mathbf{w}^T\mathbf{x}, \quad \mathbf{x}\in D_i, \quad i\in\{1,2\}$ and $\tilde{s}_i^2 = \sum_{y\in Y_i} (y-\tilde{m}_i)^2$

$$\begin{aligned}\tilde{s}_i^2 &= \sum_{\mathbf{x}\in D_i} (\mathbf{w}^T\mathbf{x} - \mathbf{w}^T\mathbf{m}_i)^2 \\ &= \mathbf{w}^T\mathbf{S}_i\mathbf{w} \\ &= \sum_{\mathbf{x}\in D_i} \mathbf{w}^T(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T\mathbf{w}\end{aligned}$$

Fisher's Discriminant for two category case

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_\mathbf{w} \mathbf{w}}$$ is always a scalar quantity

$$\mathbf{S}_B = f(\mathbf{w})\mathbf{S}_\mathbf{w}$$ must hold for a scalar valued function *f* of a vector variable **w**, because $\mathbf{w}^\mathbf{T}(\mathbf{S}_\mathrm{B} - \mathrm{f}(\mathbf{w})\,\mathbf{S}_\mathbf{w})\mathbf{w} = 0$.

Clearly, maximum *f*(**w**) will make *J*(**w**) maximum. Let maximum *f*(**w**) $= \lambda$. Then we can write

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_\mathbf{w} \mathbf{w}$$ where **w** is the vector for which *J*(**w**) is maximum.

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_{\mathbf{w}} \mathbf{w}}$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{a}}{\mathbf{w}^T \mathbf{b}}$$ **where** $$\mathbf{a} = \mathbf{S}_B \mathbf{w}$$ $$\mathbf{b} = \mathbf{S}_W \mathbf{w}$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = (\mathbf{w}^T \mathbf{a})\mathbf{b} - (\mathbf{w}^T \mathbf{b})\mathbf{a} = 0$$

For maximum **w**

$$\mathbf{a} = \frac{(\mathbf{w}^T \mathbf{a})}{(\mathbf{w}^T \mathbf{b})} \mathbf{b}$$

$$= \frac{(\mathbf{w}^T \mathbf{a})}{(\mathbf{w}^T \mathbf{b})} \mathbf{b} = J(\mathbf{w})\mathbf{b} = \lambda \mathbf{b}$$

$$\mathbf{S}_B\mathbf{w} = \lambda\mathbf{S}_\mathbf{w}\mathbf{w}$$

$$\mathbf{S}_W{}^{-1}\mathbf{S}_B\mathbf{w} = \lambda\mathbf{w}$$

Assuming $\mathbf{S}_W$ is full rank.

$$J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}} = \frac{\mathbf{w}^T\lambda\mathbf{S}_W\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}} = \lambda(\frac{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}) = \lambda$$

$$\max_\mathbf{w} J(\mathbf{w}) = \max\lambda$$

We project onto the eigen vector corresponding to the largest eigen value of $\mathbf{S}_W{}^{-1}\mathbf{S}_B$

Fisher's Discriminant for two category case

- $\mathbf{S}_B\mathbf{w}$ is in direction of $\mathbf{m}_1 - \mathbf{m}_2$. Also scale of $\mathbf{w}$ does not matter, only direction does. So we can write

$$\mathbf{S}_B\mathbf{w} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w}$$
$$= (\mathbf{m}_1 - \mathbf{m}_2)\{(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w}\}$$

This implies $\mathbf{S}_B\mathbf{w}$ is in the direction of $\mathbf{m}_1 - \mathbf{m}_2$.

Multiple Discriminant Analysis

- Fisher Discriminant for *C* Category case.

Within-Class Scatter Matrix

Fisher Discriminant for C **classes**

Assume $d > C$

$$\mathbf{S}_W = \sum_{i=1}^{c} \mathbf{S}_i$$

$$\mathbf{S}_i = \sum_{x \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{x \in D_i} \mathbf{x}$$

Total mean vector

n : Number of training samples across all classes

n_i : Number of training samples in class ω_i

Total Mean Vector

$$\mathbf{m} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x} = \frac{1}{n}\sum_{i=1}^{c} n_i \mathbf{m}_i$$

where $$n = \sum_{i=1}^{c} n_i$$

Total Scatter Matrix

$$\mathbf{S}_T = \sum_{x} (\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T$$

$$= \sum_{i=1}^{c} \sum_{x \in D_i} (\mathbf{x} - \mathbf{m}_i + \mathbf{m}_i - \mathbf{m})(x - \mathbf{m}_i + \mathbf{m}_i - \mathbf{m})^T$$

$$= \sum_{i=1}^{c} \sum_{x \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T + \sum_{i=1}^{c} \sum_{x \in D_i} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$$

$$= \mathbf{S}_W + \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$$

Between class scatter

$$\mathbf{S}_B = \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$$

Rank = $c-1$

Sum of c **rank 1 matrices gives a rank of** c**. However , only** $c-1$ **matrices are independent, owing to constraint:**

$$\mathbf{m} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x} = \frac{1}{n}\sum_{i=1}^{c} n_i \mathbf{m}_i$$

Hence, rank = $c-1$

Total scatter

$$\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B$$

Total scatter Matrix = Within class scatter matrix + Between class scatter matrix

Basic Idea

$$y_i = \mathbf{w}_i^T \mathbf{x} \qquad i = 1, \dots, c-1$$

$$\mathbf{y} = \mathbf{W}^T \mathbf{x}$$

Projection of a feature vector to a lower dimension

$$\tilde{\mathbf{m}}_i = \frac{1}{n_i} \sum_{y \in Y_i} \mathbf{y}$$

Projection of the class mean vector in lower dimension

$$\tilde{\mathbf{m}} = \frac{1}{n} \sum_{i=1}^{c} n_i \tilde{\mathbf{m}}_i$$

Projection of pooled mean vector to a lower dimension

$$\tilde{\mathbf{S}}_W = \sum_{i=1}^{c} \sum_{\mathbf{y} \in Y_i} (\mathbf{y} - \tilde{\mathbf{m}}_i)(\mathbf{y} - \tilde{\mathbf{m}}_i)^T$$

$$\tilde{\mathbf{S}}_W = \mathbf{W}^T \mathbf{S}_W \mathbf{W}$$

Within class Scatter matrix expression after projection to lower dimension

$$\tilde{\mathbf{S}}_B = \sum_{i=1}^{c} n_i (\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})(\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})^T$$

$$\tilde{\mathbf{S}}_B = \mathbf{W}^T \mathbf{S}_B \mathbf{W}$$

Between class Scatter matrix expression after projection to lower dimension

$$J(\mathbf{W}) = \frac{|\tilde{\mathbf{S}}_B|}{|\tilde{\mathbf{S}}_W|} = \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_W \mathbf{W}|}$$

Fisher Criterion function that needs to be maximized

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_w \mathbf{w}_i$$

$$(\mathbf{S}_B - \lambda_i \mathbf{S}_w)\mathbf{w}_i = 0$$

The expression for the weight vector

$$|\mathbf{S}_B - \lambda_i \mathbf{S}_w| = 0$$