The vector a in Fisher's linear discriminant analysis

December 3, 2008

Theorem 0.1 *Let $B(p \times p)$ and $W(p \times p)$ be two symmetric matrices. Then the vector a that maximizes $(a'Ba)/(a'Wa)$ is given by the eigenvector of $W^{-1}B$ corresponding to the largest eigenvalue (see Th. 11.5.1 and Th. A.9.2 of Mardia, Kent and Bibby).*

Proof: Since the vector a can be re-scaled arbitrarily without affecting the ratio $(a'Ba)/(a'Wa)$, we can reformulate the problem as follows: We need to find the vector a that maximizes $a'Ba$ subject to the constraint that $a'Wa = 1$.

Let $W^{1/2}$ be the symmetric square root of W. Let $z = W^{1/2}a$ (so that $a = W^{-1/2}z$). Then

$$a'Ba = (W^{-1/2}z)'B(W^{-1/2}z) = z'W^{-1/2}BW^{-1/2}z$$

and

$$a'Wa = (W^{-1/2}z)'W(W^{-1/2}z) = z'z.$$

Hence, the maximum of $a'Ba$ subject to the constraint that $a'Wa = 1$ is the same as the maximum of $z'W^{-1/2}BW^{-1/2}z$ subject to the constraint that $z'z = 1$.

We now consider the matrix $W^{-1/2}BW^{-1/2}$. This is a symmetric matrix, and hence we can form its spectral decomposition $\Gamma\Lambda\Gamma'$, where Λ is a diagonal matrix containing the eigenvalues of $W^{-1/2}BW^{-1/2}$ (in descending order) and $\Gamma = [\gamma_{(1)}, \dots, \gamma_{(p)}]$ is a matrix whose columns consist of the corresponding standardized eigenvectors. Now let $w = \Gamma'z$ (so that $z = \Gamma w$, since $\Gamma\Gamma' = I$). Then

$$z'W^{-1/2}BW^{-1/2}z = z'\Gamma\Lambda\Gamma'z = w'\Lambda w$$

and

$$z'z = z'\Gamma\Gamma'z = w'w.$$

So we can reformulate the problem once again: We need to find the vector w that maximizes $w'\Lambda w = \sum_{i=1}^{p} \lambda_i w_i^2$ subject to the constraint that $w'w = 1$.

Since $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, we have $\max_w \sum_{i=1}^{p} \lambda_i w_i^2 \leq \lambda_1 \max_w \sum_{i=1}^{p} w_i^2$, and this equals λ_1 for all vectors w satisfying $w'w = 1$. Moreover, the bound λ_1 is attained exactly for $w = (1, 0, \ldots, 0)$. Hence, the vector w that maximizes $w'\Lambda w$ subject to the constraint that $w'w = 1$ is $w = (1, 0, \ldots, 0)$. In turn, this means that $z = \Gamma w = \gamma_{(1)}$, and $a = W^{-1/2}z = W^{-1/2}\gamma_{(1)}$.

Finally, we use the fact that for any two matrices $A(n \times p)$ and $C(p \times n)$, the non-zero eigenvalues of AC and CA are the same and have the same multiplicity (Th. A.6.2 of Mardia, Kent and Bibby). Taking $A = W^{-1/2}B$ and $C = W^{-1/2}$, this means that the non-zero eigenvalues of $AC = W^{-1/2}BW^{-1/2}$ are the same as the non-zero eigenvalues of $CA = W^{-1}B$. Hence, λ_1 is also the largest eigenvalue of $W^{-1}B$. Moreover, since $\gamma_{(1)}$ is the eigenvector corresponding to the largest eigenvalue λ_1 of $W^{-1/2}BW^{-1/2}$, we have that

$$\begin{aligned} W^{-1}B(W^{-1/2}\gamma_{(1)}) &= W^{-1/2}(W^{-1/2}BW^{-1/2}\gamma_{(1)}) = W^{-1/2}\lambda_1\gamma_{(1)} \\ &= \lambda_1(W^{-1/2}\gamma_{(1)}). \end{aligned}$$

This shows that $a = W^{-1/2}\gamma_{(1)}$ is the eigenvector of $W^{-1}B$ corresponding to its largest eigenvalue λ_1.