JHU Engineering for Professionals
Applied and Computational Mathematics
Data Mining: 625.740

Homework for Module 9

1. Fisher's linear discriminant is

$$\hat{\mathbf{w}}^* = \arg\max_{\hat{\mathbf{w}}} J(\hat{\mathbf{w}}) = \arg\max_{\hat{\mathbf{w}}} \frac{\hat{\mathbf{w}}^T \mathbf{S}_b \hat{\mathbf{w}}}{\hat{\mathbf{w}}^T \mathbf{S}_w \hat{\mathbf{w}}},$$

where $\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$ and $\mathbf{S}_w = \sum_j \sum_\alpha (\mathbf{x}_\alpha - \mathbf{m}_j)(\mathbf{x}_\alpha - \mathbf{m}_j)^T$.

   (a) By writing $\frac{\partial J}{\partial \hat{\mathbf{w}}} = 0$, show that

$$\mathbf{S}_w^{-1} \mathbf{S}_b \hat{\mathbf{w}} = J(\hat{\mathbf{w}}) \hat{\mathbf{w}}.$$

   (b) Explain why $\hat{\mathbf{w}}^*$ is the eigenvector for which $J(\hat{\mathbf{w}})$ is the maximum eigenvalue of $\mathbf{S}_w^{-1}\mathbf{S}_b$.

   (c) Explain why $\mathbf{S}_b\hat{\mathbf{w}}$ is always in the direction of $\mathbf{m_1} - \mathbf{m_2}$ and thus show that

$$\hat{\mathbf{w}}^* \;=\; \text{const.} \cdot \mathbf{S}_w^{-1} \, (\mathbf{m_1} - \mathbf{m_2}).$$

2. Another way to optimize Fisher's linear discriminant (suggested by Barry Fridling):

   (a) Show that for any two real vectors $\mathbf{x}$ and $\mathbf{y}$

$$(\mathbf{x}^T\mathbf{y})^2 \leq (\mathbf{x}^T\mathbf{x})(\mathbf{y}^T\mathbf{y}), \quad \text{(Cauchy-Schwarz)}.$$

   (b) Show that if the λ_k are positive,

$$\left(\sum_{k=1}^{N} x_k y_k\right)^2 \leq \left(\sum_{k=1}^{N} \lambda_k x_k^2\right)\left(\sum_{k=1}^{N} y_k^2/\lambda_k\right).$$

   (c) Thus, show that for $\mathbf{A}$ positive definite

$$(\mathbf{x}^T\mathbf{y})^2 \leq (\mathbf{x}^T\mathbf{A}\mathbf{x})(\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}).$$

   (d) By letting $\mathbf{A} = \mathbf{S_w}$ in the expression above, and writing

$$J(\hat{\mathbf{w}}) = \frac{|\hat{\mathbf{w}}^T(\mathbf{m_1} - \mathbf{m_2})|^2}{\hat{\mathbf{w}}^T\mathbf{S}_w\hat{\mathbf{w}}},$$

   show again that

$$\hat{\mathbf{w}}^* \;=\; \text{const.} \cdot \mathbf{S}_w^{-1} \, (\mathbf{m_1} - \mathbf{m_2}).$$

3. Using the Optdigits dataset from the UCI repository, implement PCA. Reconstruct the digit images and calculate the reconstruction error $E(n) = \sum_j ||\hat{\mathbf{x}}_j - \mathbf{x}||^2$ for various values of n, the number of eigenvectors. Plot $E(n)$ versus n.