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

Module 9 Homework Solutions (part I)

1. $J(\hat{\mathbf{w}}) = \dfrac{\hat{\mathbf{w}}^T \mathbf{S}_b \hat{\mathbf{w}}}{\hat{\mathbf{w}}^T \mathbf{S}_w \hat{\mathbf{w}}}$, 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) $0 = \frac{1}{2}\frac{\partial J}{\partial \hat{\mathbf{w}}} = \mathbf{S}_b \hat{\mathbf{w}} - J(\hat{\mathbf{w}})\mathbf{S}_w \hat{\mathbf{w}} \Rightarrow \mathbf{S}_w^{-1}\mathbf{S}_b \hat{\mathbf{w}} = J(\hat{\mathbf{w}})\hat{\mathbf{w}}$.

   (b) $\frac{\partial J}{\partial \hat{\mathbf{w}}} = 0$ implies that $\hat{\mathbf{w}}$ is a stationary point of $J(\hat{\mathbf{w}})$. If a maximum exists, it will be one of the stationary points. Since, in equation (a), $J(\hat{\mathbf{w}})$ is an eigenvalue of $\mathbf{S}_w^{-1}\mathbf{S}_b$, and by definition $\hat{\mathbf{w}}^*$ maximizes $J(\hat{\mathbf{w}})$, $J(\hat{\mathbf{w}}^*)$ is the maximum eigenvalue of $\mathbf{S}_w^{-1}\mathbf{S}_b$ and $\hat{\mathbf{w}}^*$ is the eigenvector associated with this eigenvalue.

   (c) Since matrix multiplication is associative, $\mathbf{S}_b \hat{\mathbf{w}} = (\mathbf{m}_1 - \mathbf{m}_2)[(\mathbf{m}_1 - \mathbf{m}_2)^T \hat{\mathbf{w}}]$, which is in the direction of $(\mathbf{m}_1 - \mathbf{m}_2)$. Thus, $\hat{\mathbf{w}}^* = \frac{1}{J(\hat{\mathbf{w}}^*)}\mathbf{S}_w^{-1}\mathbf{S}_b \hat{\mathbf{w}}^* = \frac{(\mathbf{m}_1 - \mathbf{m}_2)^T \hat{\mathbf{w}}^*}{J(\hat{\mathbf{w}}^*)}\mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$, so that $\hat{\mathbf{w}}^* = \text{const.} \cdot \mathbf{S}_w^{-1} .(\mathbf{m}_1 - \mathbf{m}_2)$.

2. (a) Consider $f(\alpha) = (\mathbf{x} - \alpha\mathbf{y})^T(\mathbf{x} - \alpha\mathbf{y}) \geq 0, \quad \forall\alpha$.
   The function $f(\alpha) = \mathbf{x}^T\mathbf{x} - 2\alpha\mathbf{x}^T\mathbf{y} + \alpha^2\mathbf{y}^T\mathbf{y} \geq 0$ is quadratic in α and non-negative. Therefore, its discriminant $4(\mathbf{x}^T\mathbf{y})^2 - 4(\mathbf{x}^T\mathbf{x})(\mathbf{y}^T\mathbf{y})$ is less than or equal to zero.
   We have thus shown the Cauchy-Schwarz inequality:
   $$(\mathbf{x}^T\mathbf{y})^2 \leq (\mathbf{x}^T\mathbf{x})(\mathbf{y}^T\mathbf{y}).$$

   (b) Let $\xi_k = \sqrt{\lambda_k}x_k$ and $\eta_k = \frac{1}{\sqrt{\lambda_k}}y_k$. Then the Cauchy-Schwarz inequality
   $$(\boldsymbol{\xi}^T\boldsymbol{\eta})^2 \leq (\boldsymbol{\xi}^T\boldsymbol{\xi})(\boldsymbol{\eta}^T\boldsymbol{\eta}),$$
   implies
   $$\left(\sum_{k=1}^N x_k y_k\right)^2 \leq \left(\sum_{k=1}^N \lambda_k x_k^2\right)\left(\sum_{k=1}^N y_k^2/\lambda_k\right).$$

   (c) The matrix $\mathbf{A}$ is positive definite if and only if $\mathbf{A}_{\text{sym}} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$ is positive definite. Since $\mathbf{A}_{\text{sym}}$ is symmetric and positive definite, it has a square root which can be expressed as
   $$\mathbf{A}_{\text{sym}} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T) = \mathbf{B}\mathbf{B}^T$$
   Letting $\boldsymbol{\xi} = \mathbf{B}^T\mathbf{x}$ and $\boldsymbol{\eta} = \mathbf{B}^{-1}\mathbf{y}$,
   $$\begin{aligned}(\mathbf{x}^T\mathbf{y})^2 = (\boldsymbol{\xi}^T\boldsymbol{\eta})^2 &\leq (\boldsymbol{\xi}^T\boldsymbol{\xi})(\boldsymbol{\eta}^T\boldsymbol{\eta}) \\ &= [\mathbf{x}^T\mathbf{B}\mathbf{B}^T\mathbf{x}][\mathbf{y}^T(\mathbf{B}\mathbf{B}^T)^{-1}\mathbf{y}] \\ &= \frac{1}{4}[\mathbf{x}^T(\mathbf{A} + \mathbf{A}^T)\mathbf{x}][\mathbf{y}^T(\mathbf{A} + \mathbf{A}^T)^{-1}\mathbf{y}]\end{aligned}$$
   But $\mathbf{x}^T\mathbf{A}\mathbf{x}$ and $\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}$ are scalars, and thus equal to their respective transposes. Therefore
   $$(\mathbf{x}^T\mathbf{y})^2 \leq (\mathbf{x}^T\mathbf{A}\mathbf{x})(\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}).$$

(d) In (c), equality holds when $\boldsymbol{\xi}$ is a constant multiple of $\boldsymbol{\eta}$.
Letting $\mathbf{x} = \hat{\mathbf{w}}^*$, $\mathbf{y} = \mathbf{m}_1 - \mathbf{m}_2$, and $\mathbf{A} = \mathbf{S}_w$, with $\boldsymbol{\xi} = \text{const.} \cdot \boldsymbol{\eta}$,

$$\begin{aligned}
\mathbf{B}^T\mathbf{x} &= \text{const.} \cdot \mathbf{B}^{-1}\mathbf{y} \\
\mathbf{x} &= \text{const.} \cdot \mathbf{A}^{-1}\mathbf{y} \\
\hat{\mathbf{w}}^* &= \text{const.} \cdot \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)
\end{aligned}$$

Notice that $\mathbf{S}_w$ is positive definite and symmetric, so $\mathbf{A} = \mathbf{S}_w = \mathbf{A}^T$ and we can express $\mathbf{A}$ as $\mathbf{A} = \mathbf{B}\mathbf{B}^T$ and thus $\mathbf{A}^{-1} = \mathbf{B}^{-T}\mathbf{B}^{-1}$.