Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Exercise 9
- Proposed Solution -

Friday, December 21, 2018

Solution of Problem 1

Note that the discriminant rule is to allocate $\mathbf{x}$ to the group 1 if $|\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_1| < |\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_2|$ with $\mathbf{a} = \mathbf{W}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)$. See that:

$$\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) = \mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) + \mathbf{a}^T(\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1),$$

and note that since $\mathbf{W}^{-1}$ is nonnegative definite, we have:

$$\mathbf{a}^T(\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1) = (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T\mathbf{W}^{-1}(\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1) \leq 0,$$

hence $\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) \leq \mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2)$. We have three cases:

- If $\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) > 0$, then $|\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_1| < |\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_2|$, and the discriminant rule implies that $\mathbf{x}$ is allocated to C_1.
- If $\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) < 0$, then $|\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_1| > |\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_2|$, and the discriminant rule implies that $\mathbf{x}$ is allocated to C_2.
- If $\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) > 0$ and $\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) < 0$, the discriminant rule implies that $\mathbf{x}$ is allocated to C_1 if :
$$\mathbf{a}^T(-\mathbf{x} + \overline{\mathbf{x}}_1) < \mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) \implies \mathbf{a}^T(2\mathbf{x} - \overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1) > 0$$

Now just see that if $\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) > 0$, then $\mathbf{a}^T(2\mathbf{x} - \overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1) > 0$. And if $\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) < 0$, then $\mathbf{a}^T(2\mathbf{x} - \overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1) < 0$.

Another solution:
First of all, the discriminant rule can be simplified as follows:

$$\begin{aligned}
&|\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_1| < |\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_2| \implies \\
&(\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_1)^2 < (\mathbf{a}^T\mathbf{x} - \mathbf{a}^T\overline{\mathbf{x}}_2)^2 \implies \\
&(\mathbf{x} - \overline{\mathbf{x}}_1)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) < (\mathbf{x} - \overline{\mathbf{x}}_2)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2).
\end{aligned}$$

Note that:

$$\begin{aligned}
(\mathbf{x} - \overline{\mathbf{x}}_1)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) &= (\mathbf{x} - \overline{\mathbf{x}}_2 + \overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2 + \overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1) \\
&= (\mathbf{x} - \overline{\mathbf{x}}_2)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) + (\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) \\
&+ (\mathbf{x} - \overline{\mathbf{x}}_2)^T\mathbf{a}\mathbf{a}^T(\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1) \\
&= (\mathbf{x} - \overline{\mathbf{x}}_2)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) + (\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_1) \\
&+ (\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) \\
&= (\mathbf{x} - \overline{\mathbf{x}}_2)^T\mathbf{a}\mathbf{a}^T(\mathbf{x} - \overline{\mathbf{x}}_2) - (\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)^T\mathbf{a}\mathbf{a}^T(2\mathbf{x} - \overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)
\end{aligned}$$

Using this equlity in the discriminant rule, we obtain the rule as:

$$(\overline{\mathbf{x}}_{\mathbf{1}}-\overline{\mathbf{x}}_{\mathbf{2}})^T\mathbf{a}\mathbf{a}^T(2\mathbf{x}-\overline{\mathbf{x}}_{\mathbf{1}}-\overline{\mathbf{x}}_{\mathbf{2}})>0.$$

However since $\mathbf{W}^{-1}$ is nonnegative definite (see above), $(\overline{\mathbf{x}}_{\mathbf{1}}-\overline{\mathbf{x}}_{\mathbf{2}})^T\mathbf{a}>0$ and therefore it suffices that:

$$\mathbf{a}^T(2\mathbf{x}-\overline{\mathbf{x}}_{\mathbf{1}}-\overline{\mathbf{x}}_{\mathbf{2}})>0.$$

Solution of Problem 2

The ML discriminant rule for classification into two classes C_1 and C_2 allocates $\mathbf{x}$ to C_1 if:

$$f_1(\mathbf{x})>f_2(\mathbf{x}),$$

or equivalently if:

$$(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)<(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2).$$

Note that:

$$\begin{aligned}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)&=(\mathbf{x}-\boldsymbol{\mu}_2+\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2+\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\\
&=(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)+(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\\
&+(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\\
&=(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)+(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\\
&+(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\\
&=(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)-(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(2\mathbf{x}-\boldsymbol{\mu}_1--\boldsymbol{\mu}_2)\end{aligned}$$

Using this equlity in the discriminant rule, we have:

$$(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(2\mathbf{x}-\boldsymbol{\mu}_1--\boldsymbol{\mu}_2)>0,$$

which is the desired expression.

Solution of Problem 3

Note that $\mathbf{B}=\sum_{l=1}^{g}n_l(\overline{\mathbf{x}}_{\mathbf{l}}-\overline{\mathbf{x}})(\overline{\mathbf{x}}_{\mathbf{l}}-\overline{\mathbf{x}})^T$ and $\mathbf{W}=\sum_{l=1}^{g}\mathbf{X}_l^T\mathbf{E}_l\mathbf{X}_l$. But the crucial identity for this problem is the followin:

$$\mathbf{S}=\mathbf{B}+\mathbf{W}.$$

First of all, let $(\lambda,\mathbf{v})$ be eigenvalue-eigenvector pair for the matrix $\mathbf{W}^{-1}\mathbf{B}$. We have:

$$\mathbf{W}^{-1}\mathbf{S}=\mathbf{W}^{-1}\mathbf{B}+\mathbf{I}\implies\mathbf{W}^{-1}\mathbf{S}\mathbf{v}=\mathbf{W}^{-1}\mathbf{B}\mathbf{v}+\mathbf{v}=(\lambda+1)\mathbf{v}.$$

Therefore $(\lambda+1,\mathbf{v})$ is an eigenvalue-eigenvector pair for $\mathbf{W}^{-1}\mathbf{S}$. Moreover it can be seen that

$$\mathbf{W}^{-1}\mathbf{S}\mathbf{v}=(\lambda+1)\mathbf{v}\implies v=(\lambda+1)\mathbf{S}^{-1}\mathbf{W}\mathbf{v},$$

which means that $(\frac{1}{\lambda+1},\mathbf{v})$ is an eigenvalue-eigenvector pair for $\mathbf{S}^{-1}\mathbf{W}$. Therefore the equivalence of three eigenvectors follow these discussions.