Machine Learning 1 EX 05

23.11.2015

Group QXNXXJ

1. **Analytic Fisher**

a) calculate the derivative of $\mathbf{J}(\mathbf{w})$:

$$\frac{\partial \mathbf{J}(\mathbf{w})}{\partial \mathbf{w}} = \frac{2\mathbf{S_B}\mathbf{w}(\mathbf{w^T}\mathbf{S_w}\mathbf{w}) - 2\mathbf{S_w}\mathbf{w}(\mathbf{w^T}\mathbf{S_B}\mathbf{w})}{\mathbf{w^T}\mathbf{S_w}\mathbf{w}\mathbf{w^T}\mathbf{S_w}\mathbf{w}}$$

set this to 0, we get:

$$\begin{aligned}\mathbf{S_B}\mathbf{w}(\mathbf{w^T}\mathbf{S_w}\mathbf{w}) - \mathbf{S_w}\mathbf{w}(\mathbf{w^T}\mathbf{S_B}\mathbf{w}) &= 0\\ \mathbf{S_B}\mathbf{w}(\mathbf{w^T}\mathbf{S_w}\mathbf{w}) &= \mathbf{S_w}\mathbf{w}(\mathbf{w^T}\mathbf{S_B}\mathbf{w})\\ \mathbf{S_B}\mathbf{w} &= \mathbf{S_w}\mathbf{w}\cdot\frac{\mathbf{w^T}\mathbf{S_B}\mathbf{w}}{\mathbf{w^T}\mathbf{S_w}\mathbf{w}}\end{aligned}$$

$\dfrac{\mathbf{w^T}\mathbf{S_B}\mathbf{w}}{\mathbf{w^T}\mathbf{S_w}\mathbf{w}}$ is a scalar, set it to λ. Then this can be written as:

$$\mathbf{S_B}\mathbf{w} = \lambda\mathbf{S_w}\mathbf{w} \tag{1}$$

So the vector $\mathbf{w}$ that maximizes th objective $\mathbf{J}(\mathbf{w})$ is also a solution of eigenvalue problem.

b)If $\mathbf{S_w}$ is invertible, equation (1) turns to:

$$\mathbf{S_w^{-1}}\cdot(\mathbf{S_B}\mathbf{w}) = \lambda\mathbf{w}$$

As $\mathbf{S_B} = (\mathbf{m_1} - \mathbf{m_2})(\mathbf{m_1} - \mathbf{m_2})^T$,matrix $\mathbf{S_B}$ is an outer product of a vector, it has several properties:

$\mathbf{S_B}$ has rank 1:

$$\begin{aligned}\mathbf{m_1} - \mathbf{m_2} &= (a_1, a_2, \cdots, a_n)^T \triangleq \mathbf{a}\\ \mathbf{S_B} = (\mathbf{m_1} - \mathbf{m_2})(\mathbf{m_1} - \mathbf{m_2})^T &= (a_1\mathbf{a^T}, a_2\mathbf{a^T}, \cdots, a_n\mathbf{a^T})^T\\ \mathbf{S_B}\mathbf{w} = (a_1\mathbf{a^T}\mathbf{w}, a_2\mathbf{a^T}\mathbf{w}, \cdots, a_n\mathbf{a^T}\mathbf{w})^T &= \mathbf{a^T}\mathbf{w}(a_1, a_2, \cdots, a_n)^T = (\mathbf{a^T}\mathbf{w})\mathbf{a}\end{aligned}$$

So $\mathbf{S_B}\mathbf{w}$ is always in the same direction as $(\mathbf{m_1} - \mathbf{m_2})$

Since $\|\mathbf{w}\|$ is not important, we just need th direction. Finally we get:

$$\mathbf{w} = \mathbf{S_w^{-1}}(\mathbf{m_1} - \mathbf{m_2}) \tag{2}$$

2. **Fisher and Bayes**

a)

$$\mathbf{\Sigma_1} = \mathbf{\Sigma_2} \qquad \Rightarrow \qquad \mathbf{\Sigma_w} = \mathbf{\Sigma_1} = \mathbf{\Sigma_2} \triangleq \mathbf{\Sigma}$$

According to equation(2) from Exercise 1, $\mathbf{w}^*$ that maximizes $\mathbf{J_{Fisher}}(\mathbf{w})$ is:

$$\mathbf{w}^* = \mathbf{\Sigma^{-1}}(\mu_\mathbf{1} - \mu_\mathbf{2})$$

According to the assumption:

$$\begin{aligned} & p(\mathbf{x}|\omega_1) \sim \mathcal{N}(\mu_\mathbf{1}, \mathbf{\Sigma}), p(\mathbf{x}|\omega_2) \sim \mathcal{N}(\mu_\mathbf{2}, \mathbf{\Sigma}) \\ \Rightarrow \qquad & p(\mathbf{w^T x}|\omega_1) \sim \mathcal{N}(\mathbf{w^T}\mu_\mathbf{1}, \mathbf{w^T \Sigma w}), p(\mathbf{w^T x}|\omega_2) \sim \mathcal{N}(\mathbf{w^T}\mu_\mathbf{2}, \mathbf{w^T \Sigma w}) \end{aligned}$$

Now we know the distributions of $p(\mathbf{w^T x}|\omega_1)$ and $p(\mathbf{w^T x}|\omega_2)$

$$\begin{aligned} \mathbf{J_{Bayes}}(\mathbf{w}) &= \int \min\{\mathrm{P}(\omega_1|\mathbf{w^T x}), \mathrm{P}(\omega_2|\mathbf{w^T x})\} p(\mathbf{x}) d\mathbf{x} \\ &= \int_{\mathbf{R_1}} \mathrm{P}(\omega_2|\mathbf{w^T x}) p(\mathbf{x}) d\mathbf{x} + \int_{\mathbf{R_2}} \mathrm{P}(\omega_1|\mathbf{w^T x}) p(\mathbf{x}) d\mathbf{x} \end{aligned}$$

$\mathbf{R_i}$ is the region where a Bayes classificator decides class $\omega_i \qquad i = 1, 2$
Apply Bayes formula:

$$\mathrm{P}(\omega_i|\mathbf{w^T x}) = \frac{\mathrm{P}(\mathbf{w^T x}|\omega_i)\mathbb{P}(\omega_i)}{\mathrm{P}(\mathbf{w^T x})}$$

In this way, $\mathbf{J_{Bayes}}(\mathbf{w})$ can be written as a function of $\mathbf{w}$, calculate the derivative and set it to 0, we get the same answer as $\mathbf{w}^*$.
b)When these two distributions have different shape, Fisher-method may have a different solution from Bayes-method.