Modern Multivariate Statistical Techniques

Jinhong Du,15338039

April 22, 2018

Content

1. Suppose $\mathbb{E}\mathbf{X}_1 = \mu_1$, $Var\mathbf{X}_1 = \Sigma_{XX}$ and $\mathbb{E}\mathbf{X}_2 = \mu_2$, $Var\mathbf{X}_2$ are independently distributed. Consider the statistic
$$\frac{\{\mathbb{E}(\mathbf{a}^\top\mathbf{X}_1) - \mathbb{E}(\mathbf{a}^\top\mathbf{X}_2)\}^2}{Var(\mathbf{a}^\top\mathbf{X}_1 - \mathbf{a}^\top\mathbf{X}_2)}$$
as a function of $\mathbf{a}$. Show that $\mathbf{a} \propto \Sigma_{XX}^{-1}(\mu_1 - \mu_2)$ maximizes the statistic by using a Lagrange multiplier approach.

Proof.

Let $\mathbf{a}^\top\Sigma_{XX}\mathbf{a} = 1$, then
$$L(\mathbf{a}, \lambda) = \mathbf{a}^\top(\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top\mathbf{a} - \lambda(\mathbf{a}^\top\Sigma_{XX}\mathbf{a} - 1)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 2(\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top\mathbf{a} - 2\lambda\Sigma_{XX}\mathbf{a} = 0 \quad (1)$$
$$\frac{\partial L}{\partial \lambda} = -\mathbf{a}^\top\Sigma_{XX}\mathbf{a} + 1 = 0 \quad (2)$$

The maximizor should satisfy
$$(\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top\mathbf{a}^* = \lambda\Sigma_{XX}\mathbf{a}^*$$

Since $(\mu_1 - \mu_2)^\top\mathbf{a}^* = c$ is a scalar, so we can write as
$$c(\mu_1 - \mu_2) = \lambda\Sigma_{XX}\mathbf{a}^*$$

And therefore,
$$\begin{aligned}\mathbf{a}^* &= \frac{c}{\lambda}\Sigma_{XX}^{-1}(\mu_1 - \mu_2)\\ &\propto \Sigma_{XX}^{-1}(\mu_1 - \mu_2)\end{aligned}$$

□

2. Consider the problem of finding $\mathbf{\Theta}^*$ that solves the following constrained minimization problem:
$$\hat{\mathbf{\Theta}}^* = \underset{\substack{\mathbf{\Theta}\\ \mathbf{K\Theta L}=\Gamma}}{\arg\min}\ tr\{(\mathbf{Y}_c - \mathbf{\Theta}\mathbf{X}_c)^\top(\mathbf{Y}_c - \mathbf{\Theta}\mathbf{X}_c)\}$$
Let $\Lambda = (\lambda_{ij})$ be a matrix of Lagrangian coefficients. The normal equations are:
$$\hat{\mathbf{\Theta}}^*\mathbf{X}_c\mathbf{X}_c^\top + \mathbf{K}^\top\Lambda\mathbf{L}^\top = \mathbf{Y}_c\mathbf{X}_c^\top$$

$$\mathbf{K}\hat{\Theta}^*\mathbf{L} = \Gamma$$

where

$$\mathbf{X}_c = \begin{pmatrix} \mathbf{X}_1 - \overline{\mathbf{X}} & \mathbf{X}_2 - \overline{\mathbf{X}} & \cdots & \mathbf{X}_n - \overline{\mathbf{X}} \end{pmatrix}$$
$$\mathbf{Y}_c = \begin{pmatrix} \mathbf{Y}_1 - \overline{\mathbf{Y}} & \mathbf{Y}_2 - \overline{\mathbf{Y}} & \cdots & \mathbf{Y}_n - \overline{\mathbf{Y}} \end{pmatrix}$$

Proof.

$$f(\Theta, \Lambda) = tr\{(\mathbf{Y}_c - \Theta\mathbf{X}_c)^\top(\mathbf{Y}_c - \Theta\mathbf{X}_c)\} - 2tr\{\Lambda^\top(\mathbf{K}\Theta\mathbf{L} - \Gamma)\}$$

Since

$$\frac{\partial tr\{F(\mathbf{X})\}}{\partial \mathbf{X}} = F_X(\mathbf{X})^\top$$
$$\frac{\partial tr\{\mathbf{X}^\top\mathbf{A}\}}{\partial \mathbf{X}} = \mathbf{A}$$
$$\frac{\partial \mathbf{a}^\top\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^\top$$
$$\frac{\partial \mathbf{b}^\top\mathbf{X}^\top\mathbf{X}\mathbf{c}}{\partial \mathbf{X}} = \mathbf{X}(\mathbf{b}\mathbf{c}^\top + \mathbf{c}\mathbf{b}^\top)$$

Let

$$\frac{\partial f}{\partial \Theta} = 2\mathbf{Y}_c\mathbf{X}_c^\top - 2\hat{\Theta}\mathbf{X}_c\mathbf{X}_c^\top - 2\mathbf{K}^\top\Lambda\mathbf{L}^\top = 0$$
$$\frac{\partial f}{\partial \Lambda} = \mathbf{K}\Theta\mathbf{L} - \Gamma = 0$$

Therefore,

$$\hat{\Theta}^*\mathbf{X}_c\mathbf{X}_c^\top + \mathbf{K}^\top\Lambda\mathbf{L}^\top = \mathbf{Y}_c\mathbf{X}_c^\top$$
$$\mathbf{K}\hat{\Theta}^*\mathbf{L} = \Gamma$$

□