Tsinghua-Berkeley Shenzhen Institute
Inference and Information
Fall 2017

Homework 2

赵丰 November 26, 2017

- **Acknowledgments:** This coursework referes to wikipedia: `https://en.wikipedia.org`.
- **Collaborators:** I finish this coursework by myself.

I use `enumerate` to generate answers for each question:

2.1. (a) i. 对向量值随机变量证明全（协）方差公式。用到 $\mathbb{E}[\underline{x}] = \mathbb{E}[\mathbb{E}[\underline{x}|y]]$ 以及 $\mathbb{E}[\underline{x}\underline{x}^{\mathrm{T}}] = \mathbb{E}[\mathbb{E}[\underline{x}|y]]$

$$\begin{aligned}\mathrm{Cov}(\underline{x}) &= \mathbb{E}[(\underline{x} - \mathbb{E}[\underline{x}])(\underline{x} - \mathbb{E}[\underline{x}])^{\mathrm{T}}] \\ &= \mathbb{E}[\underline{x}\underline{x}^{\mathrm{T}}] - \mathbb{E}[\underline{x}]\,\mathbb{E}[\underline{x}]^{\mathrm{T}} \\ &= \mathbb{E}[\mathbb{E}[\underline{x}\underline{x}^{\mathrm{T}}|y]] - \mathbb{E}[\mathbb{E}[\underline{x}|y]]\,\mathbb{E}[\mathbb{E}[\underline{x}|y]]^{\mathrm{T}} \\ &= \mathbb{E}[\mathrm{Cov}[\underline{x}|y] + \mathbb{E}[\underline{x}|y]\,\mathbb{E}[\underline{x}|y]^{\mathrm{T}}] - \mathbb{E}[\mathbb{E}[\underline{x}|y]]\,\mathbb{E}[\mathbb{E}[\underline{x}|y]]^{\mathrm{T}} \\ &= \mathbb{E}[\mathrm{Cov}[\underline{x}|y]] + \mathrm{Cov}[\mathbb{E}[\underline{x}|y]]\end{aligned}$$

所以

$$\mathbb{E}[x|y] = \mathbb{E}[\mathbb{E}[x|yz]|y] \tag{1}$$

ii. 若 $\exists \underline{c} \in \mathbb{R}^k, \underline{c} \neq \underline{0}, d \in \mathbb{R}$ 使得 $\underline{c}^{\mathrm{T}}\underline{x} = d$. 则 $\underline{c}^{\mathrm{T}}\,\mathbb{E}[\underline{x}] = d$ 且有 $\underline{c}^{\mathrm{T}}\underline{x}\underline{x}^{\mathrm{T}}\underline{c} = d^2$，取期望得 $\underline{c}^{\mathrm{T}}\,\mathbb{E}[\underline{x}\underline{x}^{\mathrm{T}}]\underline{c} = d^2$

$$\begin{aligned}\underline{c}^{\mathrm{T}}\,\mathrm{Cov}(\underline{x})\underline{c} =& \underline{c}^{\mathrm{T}}\,\mathbb{E}[\underline{x}\underline{x}^{\mathrm{T}}]\underline{c} - \underline{c}^{\mathrm{T}}\,\mathbb{E}[\underline{x}]\,\mathbb{E}[\underline{x}]^{\mathrm{T}}\underline{c} \\ =& \underline{c}^{\mathrm{T}}\,\mathbb{E}[\underline{x}\underline{x}^{\mathrm{T}}]\underline{c} - \underline{c}^{\mathrm{T}}\,\mathbb{E}[\underline{x}]\,\mathbb{E}[\underline{x}]^{\mathrm{T}}\underline{c} \\ =& 0\end{aligned}$$

因为 $\mathrm{Cov}(\underline{x})$ 是半正定矩阵，从上式知其有一个特征值为零。从而其行列式为 0。
反之，若 $\det[\mathrm{Cov}(\underline{x})] = 0$, 则 $\exists \underline{c} \in \mathbb{R}^k, \underline{c} \neq \underline{0}$，使得 $\underline{c}^{\mathrm{T}}\,\mathrm{Cov}(\underline{x})\underline{c} = 0$, 设 $y = \underline{c}^{\mathrm{T}}(\underline{x} - \mathbb{E}[\underline{x}])$ 即 $\mathbb{E}[y^2] = 0$, 从而 $y = 0$, 取 $d = \underline{c}^{\mathrm{T}}\,\mathbb{E}[\underline{x}]$, 则有 $\underline{c}^{\mathrm{T}}\underline{x} = d$

(b) i.

$$\begin{aligned}\mathrm{MSE}(\underline{\alpha}) =& \sum_{i=1}^{k}[(1-\underline{\alpha}_i)^2 + \sum_{\substack{j=1\\ j\neq i}}^{k}\underline{\alpha}_j^2]P_y(i)\\ =&1\sum_{i=1}^{k}(-2\underline{\alpha}_i + \sum_{j=1}^{k}\underline{\alpha}_j^2)P_y(i)\\ =&1-2\sum\underline{\alpha}_iP_y(i) + \sum_{j=1}^{k}(\sum_{i=1}^{k}P_y(i))\underline{\alpha}_j^2\\ =&1-2\sum_{i=1}^{k}(\underline{\alpha}_i^2 - 2\underline{\alpha}_iP_y(i))\end{aligned}$$

根据二次函数极值的性质，当 $\underline{\alpha}_i = P_y(i)$ 时上式最小。即题中所给 $\underline{P}_y(\cdot)$ 是 MMSE 估计量。

ii. 类似上次作业，只需对给定的 $\underline{\mathsf{x}} = \underline{x}_0$ 极小化条件概率下的误差期望:

$$\mathbb{E}_{\mathsf{y}|\underline{\mathsf{x}}=\underline{x}_0}[||\underline{\mathsf{y}} - \underline{f}(\underline{\mathsf{x}} = \underline{x}_0)||_2^2|] \tag{2}$$

此时 $\underline{f}(\underline{\mathsf{x}} = \underline{x}_0)$ 视为常数，由前一小问结论，当 $\underline{f}(\underline{\mathsf{x}} = \underline{x}_0) = \underline{P}_{\mathsf{y}}(\cdot|\underline{\mathsf{x}} = \underline{x}_0)$ 时上式最小，所以 MMSE 估计器是 $\underline{P}_{\mathsf{y}}(\cdot|\underline{\mathsf{x}})$

2.2. (a) $\mathrm{Var}(\underline{v}^{\mathrm{T}}\underline{\mathsf{x}}) = \underline{v}^{\mathrm{T}}\,\mathrm{Cov}(\underline{x})\underline{v}$ 设 $U^{\mathrm{T}} = (\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_d)$, 则 $\underline{u}_i$ 是 σ_i 对应的特征向量。由 Rayleigh 商的性质可说明 $\underline{v}_i = \underline{u}_i$。设 $\underline{v} = \sum_{i=1}^{d}\alpha_i\underline{u}_i$

$$\begin{aligned}\mathrm{Var}(\underline{v}^{\mathrm{T}}\underline{\mathsf{x}}) =& \frac{\underline{v}^{\mathrm{T}}\,\mathrm{Cov}(\underline{x})\underline{v}}{||\underline{v}||^2}\\ =&\frac{(\sum_{i=1}^{d}\alpha_i\underline{u}_i)^{\mathrm{T}}(\sum_{i=1}^{d}\sigma_i\alpha_i\underline{u}_i)}{(\sum_{i=1}^{d}\alpha_i\underline{u}_i)^{\mathrm{T}}(\sum_{i=1}^{d}\alpha_i\underline{u}_i)}\\ =&\frac{\sum_{i=1}^{d}\alpha_i^2\sigma_i}{\sum_{i=1}^{d}\alpha_i^2}\end{aligned} \tag{3}$$

注意到上式是 $\sigma_1, \ldots, \sigma_d$ 的凸组合，在无正交性约束的情况下取 $\alpha_i = \delta_{1i}$ 使得 σ_1 前的系数最大即得到上式的最大值，此时

$$\underline{v}_1 = \arg\max_{\underline{v}:||\underline{v}||=1}\mathrm{Var}(\underline{v}^{\mathrm{T}}\underline{\mathsf{x}}) = \underline{u}_1 \tag{4}$$

若加入正交性约束 $\underline{v} \perp \underline{u}_1$, 则 $\alpha_1 = 0$, 则在(3)式中相当于 $\sigma_2, \ldots, \sigma_d$ 的凸组合。因此取 $\alpha_i = \delta_{2i}$ 使得 σ_2 前的系数最大得到有正交性约束 $\underline{v} \perp \underline{u}_1$ 情形下的最大值。即推出 $\underline{v}_2 = \underline{u}_2$ 如此递推下去即可证明 $\underline{v}_i = \underline{u}_i$。

(b) 注意到 $\mathrm{Cov}(\mathsf{y}_i, \mathsf{y}_j) = \underline{v}_i^T\,\mathrm{Cov}(\underline{\mathsf{x}})\underline{v}_j = \sigma_j\delta_{ij}$ 因此 $\mathrm{Cov}(\underline{\mathsf{y}}) = \Sigma$。

(c) i. $\underline{\mathsf{y}}$ 也是多维正态分布，$\mathbb{E}[\underline{\mathsf{y}}] = 0$, 其联合概率密度函数可以写为：

$$p_{\underline{\mathsf{y}}}(\underline{y}) = \prod_{i=1}^{d}\frac{1}{\sqrt{2\pi\sigma_i^2}}\exp(-\frac{y_i^2}{2\sigma_i^2}) \tag{5}$$

ii. 由于 $\underline{\mathsf{y}}$ 的各分量不相关，所以彼此独立。

2.3. (a)

$$K_{\mathsf{x}} = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \tag{6}$$

$$K_{\mathsf{y}} = AK_{\mathsf{x}}A^{\mathrm{T}} \begin{bmatrix} \sigma^2 & \rho_{\mathsf{x}}\sigma^2 \\ \rho_{\mathsf{x}}\sigma^2 & \sigma^2 \end{bmatrix} = \sigma^2 \operatorname{diag}\{1-\rho_{\mathsf{x}}^2, 1\} \tag{7}$$

(b) 由 K_{y} 是对角阵以及 $\underline{\mathsf{y}}$ 服从二元正态分布可知 $\mathsf{y}_1 \perp\!\!\!\perp \mathsf{y}_2$ 所以 $\mathsf{y}_1 \perp\!\!\!\perp g(\mathsf{y}_2)$, 即 $\mathrm{Cov}(\mathsf{y}_1, g(\mathsf{y}_2)) = 0$, 从而结论得证。

(c) 为证明 $\mathbb{E}[\mathsf{y}_1^2] \leq \mathbb{E}[(\mathsf{y}_1 + \rho_x\mathsf{y}_2 - g(\mathsf{y}_2))^2]$ 因为 y_1 与 $\rho_x\mathsf{y}_2 - g(\mathsf{y}_2)$ 相互独立，故上式右端为 $\mathbb{E}[\mathsf{y}_1^2] + \mathbb{E}[(\rho_x\mathsf{y}_2 - g(\mathsf{y}_2))^2]$, 所以不等式成立。

2.4. 上机作业

2.5. Thanks to 陆石, who gives me this template.