CS446: Machine Learning, Fall 2017, Homework 4

Name: Lanxiao Bai (lbai5)

Worked individually

Problem 1

Solution:

$$
\begin{aligned}
\sum_{i=1}^{d} Var(Y_i) &= Tr(\mathbb{E}[(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^T])\\
&= \mathbb{E}[Tr((\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^T)]\\
&= \mathbb{E}[Tr((\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y}^T - \mathbb{E}[\mathbf{Y}^T]))]\\
&= \mathbb{E}[Tr((\mathbf{Y}^T - \mathbb{E}[\mathbf{Y}^T])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}]))]\\
&= \mathbb{E}[Tr(\mathbf{Y}^T\mathbf{Y} - \mathbb{E}[\mathbf{Y}]\mathbf{Y}^T - \mathbf{Y}\mathbb{E}[\mathbf{Y}^T] + \mathbb{E}[\mathbf{Y}^T]\mathbb{E}[\mathbf{Y}])]\\
&= \mathbb{E}[Tr(\mathbf{X}^T\mathbf{U}\mathbf{U}^T\mathbf{X} - \mathbb{E}[\mathbf{U}^T\mathbf{X}](\mathbf{U}^T\mathbf{X})^T - \mathbf{U}^T\mathbf{X}\mathbb{E}[(\mathbf{U}^T\mathbf{X})^T] + \mathbb{E}[(\mathbf{U}^T\mathbf{X})^T]\mathbb{E}[\mathbf{U}^T\mathbf{X}])]\\
&= \mathbb{E}[Tr(\mathbf{X}^T\mathbf{X} - \mathbb{E}[\mathbf{X}]\mathbf{X}^T - \mathbf{X}\mathbb{E}[\mathbf{X}^T] + \mathbb{E}[\mathbf{X}^T]\mathbb{E}[\mathbf{X}])]\\
&= \mathbb{E}[Tr((\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T)]\\
&= Tr(\mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T])\\
&= \sum_{i=1}^{d} Var(X_i)
\end{aligned}
\tag{1}
$$

Thus, we have

$$\sum_{i=1}^{d} \sigma_{ii} = \sum_{i=1}^{d} Var(X_i) = \sum_{i=1}^{d} \lambda_i = \sum_{i=1}^{d} Var(Y_i)$$

Problem 2

Solution:

$$\begin{aligned}
\rho_{Y_i,X_k} &= \frac{Cov(Y_i, X_k)}{\sqrt{Var(Y_i)}\sqrt{Var(X_k)}} \\
&= \frac{\mathbb{E}[Y_i \cdot X_k] - \mathbb{E}[Y_i]\mathbb{E}[X_k]}{\sqrt{\lambda_i \sigma_{kk}}} \\
&= \frac{u_{ik}\mathbb{E}[\mathbf{X}^2] - u_{ik}\mathbb{E}[\mathbf{X}]^2}{\sqrt{\lambda_i \sigma_{kk}}} \\
&= \frac{u_{ik}\Sigma}{\sqrt{\lambda_i \sigma_{kk}}} \\
&= \frac{u_{ik}\lambda_i}{\sqrt{\lambda_i \sigma_{kk}}} \\
&= \frac{u_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}
\end{aligned}$$

Problem 3

Solution: Since

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

(a) By solving

$$det(\Sigma - \lambda I) = 0 \Rightarrow \begin{bmatrix} 1-\lambda & -2 & 0 \\ -2 & 5-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{bmatrix} = 0$$

We have

$$\lambda_1 = 3 + 2\sqrt{2}$$
$$\lambda_2 = 2$$
$$\lambda_3 = 3 - 2\sqrt{2}$$

and

$$u_1 = (1 - \sqrt{2}, 1, 0)$$
$$u_2 = (0, 0, 1)$$
$$u_3 = (1 + \sqrt{2}, 1, 0)$$

(b)

$$
\begin{aligned}
Y_1 &= u_1^T X = (1-\sqrt{2})X_1 + X_2 \\
Y_2 &= u_2^T X = X_3 \\
Y_3 &= u_3^T X = (1+\sqrt{2})X_1 + X_2
\end{aligned}
$$

(c)

$$
\begin{aligned}
Var(Y_1) &= \mathbb{E}[Y_1^2] - \mathbb{E}[Y_1]^2 \\
&= \mathbb{E}[(3-2\sqrt{2})X_1^2 + 2(1-\sqrt{2})X_1X_2 + X_2^2] - \mathbb{E}[(1-\sqrt{2})X_1 + X_2]^2 \\
&= (3-2\sqrt{2})\mathbb{E}[X_1^2] + 2(1-\sqrt{2})\mathbb{E}[X_1X_2] + \mathbb{E}[X_2^2] - ((1-\sqrt{2})\mathbb{E}[X_1] + \mathbb{E}[X_2])^2 \\
&= (3-2\sqrt{2})\mathbb{E}[X_1^2] + 2(1-\sqrt{2})\mathbb{E}[X_1X_2] + \mathbb{E}[X_2^2] - \\
&((3-2\sqrt{2})\mathbb{E}[X_1]^2 + 2(1-\sqrt{2})\mathbb{E}[X_1]\mathbb{E}[X_2] + \mathbb{E}[X_2]^2) \\
&= (3-2\sqrt{2})(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) + 2(1-\sqrt{2})(\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]) + (\mathbb{E}[X_2^2] - \mathbb{E}[X_2]^2) \\
&= (3-2\sqrt{2})\sigma_{11} + 2(1-\sqrt{2})\sigma_{12} + \sigma_{22} \\
&= (3-2\sqrt{2}) - 4(1-\sqrt{2}) + 5 \\
&= 4 + 2\sqrt{2}
\end{aligned}
$$

$$
\begin{aligned}
Var(Y_2) &= \mathbb{E}[Y_2^2] - \mathbb{E}[Y_2]^2 \\
&= \mathbb{E}[X_3^2] - \mathbb{E}[X_3]^2 \\
&= \sigma_{33} \\
&= 2
\end{aligned}
$$

$$
\begin{aligned}
Var(Y_3) &= \mathbb{E}[Y_1^2] - \mathbb{E}[Y_1]^2 \\
&= \mathbb{E}[(3+2\sqrt{2})X_1^2 + 2(1+\sqrt{2})X_1X_2 + X_2^2] - \mathbb{E}[(1+\sqrt{2})X_1 + X_2]^2 \\
&= (3+2\sqrt{2})\mathbb{E}[X_1^2] + 2(1+\sqrt{2})\mathbb{E}[X_1X_2] + \mathbb{E}[X_2^2] - ((1+\sqrt{2})\mathbb{E}[X_1] + \mathbb{E}[X_2])^2 \\
&= (3+2\sqrt{2})\mathbb{E}[X_1^2] + 2(1+\sqrt{2})\mathbb{E}[X_1X_2] + \mathbb{E}[X_2^2] - \\
&((3+2\sqrt{2})\mathbb{E}[X_1]^2 + 2(1+\sqrt{2})\mathbb{E}[X_1]\mathbb{E}[X_2] + \mathbb{E}[X_2]^2) \\
&= (3+2\sqrt{2})(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) + 2(1+\sqrt{2})(\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]) + (\mathbb{E}[X_2^2] - \mathbb{E}[X_2]^2) \\
&= (3+2\sqrt{2})\sigma_{11} + 2(1+\sqrt{2})\sigma_{12} + \sigma_{22} \\
&= (3+2\sqrt{2}) - 4(1+\sqrt{2}) + 5 \\
&= 4 - 2\sqrt{2}
\end{aligned}
$$

(d)

$$
\begin{aligned}
Cov(Y_1, Y_2) &= \mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2] \\
&= \mathbb{E}[((1-\sqrt{2})X_1 + X_2)X_3] - \mathbb{E}[((1-\sqrt{2})X_1 + X_2)]\mathbb{E}[X_3] \\
&= (1-\sqrt{2})\mathbb{E}[X_1X_3] + \mathbb{E}[X_2X_3] - (1-\sqrt{2})\mathbb{E}[X_1]\mathbb{E}[X_3] - \mathbb{E}[X_2]\mathbb{E}[X_3]
\end{aligned}
$$

$$
\begin{aligned}
&= (1-\sqrt{2})(\mathbb{E}[X_1X_3] - \mathbb{E}[X_1]\mathbb{E}[X_3]) + (\mathbb{E}[X_2X_3] - \mathbb{E}[X_2]\mathbb{E}[X_3]) \\
&= (1-\sqrt{2})\sigma_{13} + \sigma_{23} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
Cov(Y_1, Y_3) &= \mathbb{E}[Y_1Y_3] - \mathbb{E}[Y_1]\mathbb{E}[Y_3] \\
&= \mathbb{E}[((1-\sqrt{2})X_1 + X_2)((1+\sqrt{2})X_1 + X_2)] - \mathbb{E}[((1-\sqrt{2})X_1 + X_2)]\mathbb{E}[(1+\sqrt{2})X_1 + X_2] \\
&= \mathbb{E}[-X_1^2 + (1-\sqrt{2})X_1X_2 + (1+\sqrt{2})X_1X_2 + X_2^2] - \\
&(((1-\sqrt{2})\mathbb{E}[X_1] + \mathbb{E}[X_2])(((1+\sqrt{2})\mathbb{E}[X_1] + \mathbb{E}[X_2]) \\
&= -(\mathbb{E}[X_1^2] - \mathbb{E}[X_1])^2) + 2(\mathbb{E}[X_1X_2] - \mathbb{E}[X_1][X_2]) + (\mathbb{E}[X_2^2] - \mathbb{E}[X_2]^2) \\
&= -\sigma_{11} + 2\sigma_{12} + \sigma_{22} \\
&= -1 - 4 + 5 \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
Cov(Y_2, Y_3) &= \mathbb{E}[Y_2Y_3] - \mathbb{E}[Y_2]\mathbb{E}[Y_3] \\
&= \mathbb{E}[((1+\sqrt{2})X_1 + X_2)X_3] - \mathbb{E}[((1+\sqrt{2})X_1 + X_2)]\mathbb{E}[X_3] \\
&= (1+\sqrt{2})\mathbb{E}[X_1X_3] + \mathbb{E}[X_2X_3] - (1+\sqrt{2})\mathbb{E}[X_2]\mathbb{E}[X_3] - \mathbb{E}[X_2]\mathbb{E}[X_3] \\
&= (1+\sqrt{2})(\mathbb{E}[X_1X_3] - \mathbb{E}[X_1]\mathbb{E}[X_3]) + (\mathbb{E}[X_2X_3] - \mathbb{E}[X_2]\mathbb{E}[X_3]) \\
&= (1+\sqrt{2})\sigma_{13} + \sigma_{23} \\
&= 0
\end{aligned}
$$

We notice the Y_1, Y_2, Y_3 are independent random variable.

(e)

$$
\sum_{i=1}^{3} \lambda_i = 3 + 2\sqrt{2} + 2 + 3 - 2\sqrt{2} = 8 = 1 + 5 + 2 = Tr(\Sigma)
$$

(f)

$$
\begin{aligned}
\textbf{Ratio}(Y_1) &= \frac{Var(Y_1)}{\sum_{i=1}^{3} Var(Y_i)} = \frac{4+2\sqrt{2}}{4+2\sqrt{2}+2+4-2\sqrt{2}} = \frac{4+2\sqrt{2}}{10} \approx 68.2\% \\
\textbf{Ratio}(Y_2) &= \frac{Var(Y_2)}{\sum_{i=1}^{3} Var(Y_i)} = \frac{2}{4+2\sqrt{2}+2+4-2\sqrt{2}} = \frac{2}{10} \approx 20\% \\
\textbf{Ratio}(Y_3) &= \frac{Var(Y_3)}{\sum_{i=1}^{3} Var(Y_i)} = \frac{4-2\sqrt{2}}{4+2\sqrt{2}+2+4-2\sqrt{2}} = \frac{4-2\sqrt{2}}{10} \approx 11.7\%
\end{aligned}
$$

Since the first two component captured about 88.2% of the total variance, so the I would choose 2 dimensions.

(g)

$$
\rho_{Y_1, X_1} = \frac{u_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{(1-\sqrt{2})(\sqrt{3+2\sqrt{2}})}{\sqrt{4+2\sqrt{2}}} \approx -0.383
$$

$$\rho_{Y_1,X_2} = \frac{u_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{\sqrt{3+2\sqrt{2}}}{\sqrt{2}} \approx 1.707$$

(h) Since we see in (b) that

$$Y_1 = (1-\sqrt{2})X_1 + X_2$$

then we see that both random variables aid in the interpretation of Y_1.