Prof. Dr. Rudolf Mathar, Dr. Gholamreza Alirezaei, Dr. Arash Behboodi

Exercise 12
- Proposed Solution -

Friday, February 10, 2017

Solution of Problem 1

Let $\mathbf{X}$ be a matrix in $\mathbb{R}^{m\times n}$ such that $(\mathbf{X}^T\mathbf{X})$ is invertible. To show that the matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is a projection matrix, we have to show $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{P}$ is symmetric. First see that:

$$\mathbf{P}^2 = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{P}.$$

For proving that P is symmetric, see that:

$$\mathbf{P}^T = \left(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right)^T = (\mathbf{X}^T)^T((\mathbf{X}^T\mathbf{X})^{-1})^T\mathbf{X}^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{P}.$$

So $\mathbf{P}$ is a projection matrix. It remain to show that $\mathbf{P}$ is the projection matrix onto the image of $\mathbf{X}$. Suppose that $\mathbf{b} \in \mathbb{R}^n$ belongs to the image of $\mathbf{X}$, therefore there is $\mathbf{a} \in \mathbb{R}^m$ such that $\mathbf{b} = \mathbf{X}\mathbf{a}$. We have:

$$\mathbf{P}\mathbf{b} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X})\mathbf{a} = \mathbf{X}\mathbf{a} = \mathbf{b}.$$

In other words every vector in the image of $\mathbf{X}$ is projected onto itself. Now note that the image of $\mathbf{P}$ is a subset of the image of $\mathbf{X}$. Therefore $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the projection matrix onto the image of $\mathbf{X}$.

Solution of Problem 2

a) Let $\mathbf{B}$ and $\mathbf{C}$ be Moore-Penrose pseudoinverses of $\mathbf{A}$. First of all see that

$$(\mathbf{BA})^T = \mathbf{BA} \implies (\mathbf{BA})^T = (\mathbf{BACA})^T = (\mathbf{CA})^T(\mathbf{BA})^T = \mathbf{CABA} = \mathbf{CA}.$$

On the other hand, we have:

$$(\mathbf{AB})^T = \mathbf{AB} \implies (\mathbf{AB})^T = (\mathbf{ACAB})^T = (\mathbf{AB})^T(\mathbf{AC})^T = \mathbf{ABAC} = \mathbf{AC}.$$

Therefore $\mathbf{CA} = \mathbf{BA}$ and $\mathbf{AB} = \mathbf{AC}$. So we have:

$$\mathbf{B}(\mathbf{AC}) = \mathbf{B}(\mathbf{AB}) = \mathbf{B}$$

and

$$(\mathbf{BA})\mathbf{C} = (\mathbf{CA})\mathbf{C} = \mathbf{C},$$

which implies that $\mathbf{B} = \mathbf{C}$.

b) Suppose that $\text{rk}(\mathbf{A}) = m$. Note that $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ and hence $\text{rk}(\mathbf{A}\mathbf{A}^T) \leq m$. On the other hand, $\text{rk}(\mathbf{A}\mathbf{A}^T) = \text{rk}(\mathbf{A}) = m$. Therefore $\mathbf{A}\mathbf{A}^T$ is full rank and invertible.

Now that $\mathbf{A}\mathbf{A}^T$ is invertible, it is enough to check the conditions of Moore-Penrose inverse:

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = (\mathbf{A}\mathbf{A}^T)(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A} = \mathbf{A}.$$

$$\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{A}^T)(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}.$$

$$(\mathbf{A}\mathbf{A}^+)^T = (\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1})^T = \mathbf{I}$$

$$(\mathbf{A}^+\mathbf{A})^T = (\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})^T = \mathbf{A}^T((\mathbf{A}\mathbf{A}^T)^{-1})^T\mathbf{A} = \mathbf{A}^+\mathbf{A}.$$

c) If $\text{rk}(\mathbf{A}) = n$, then $\text{rk}(\mathbf{A}^T\mathbf{A}) = n$ and since $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$, the matrix is full rank and invertible. Now that $(\mathbf{A}^T\mathbf{A})$ is invertible, similar to the previous exercise it can be shown that $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ satisfies Moore-Penrose condition.

d) We check all the conditions step by step:

$$\mathbf{A}\mathbf{B}\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^+\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{U}\mathbf{D}\mathbf{D}^+\mathbf{D}\mathbf{V}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{A}.$$

where we used $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ and $\mathbf{U}^T\mathbf{U} = I$ and also:

$$\mathbf{D}\mathbf{D}^+ = \begin{bmatrix}\mathbf{S} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{S}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}^T = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix} = \text{diag}(\mathbf{I}, \mathbf{0}).$$

In a similar fashion, we have:

$$\mathbf{B}\mathbf{A}\mathbf{B} = \mathbf{V}\mathbf{D}^+\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^+\mathbf{U}^T = \mathbf{V}\mathbf{D}^+\mathbf{D}\mathbf{D}^+\mathbf{U}^T = \mathbf{V}\mathbf{D}^+\mathbf{U}^T = \mathbf{B}.$$

Next step is to show that $\mathbf{B}\mathbf{A}$ and $\mathbf{A}\mathbf{B}$ are symmteric. Note that:

$$\mathbf{B}\mathbf{A} = \mathbf{V}\mathbf{D}^+\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{V}\mathbf{D}^+\mathbf{D}\mathbf{V}^T = \mathbf{V}\,\text{diag}(\mathbf{I}, \mathbf{0})\mathbf{V}^T.$$

$$\mathbf{A}\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}^+\mathbf{U}^T = \mathbf{U}\mathbf{D}\mathbf{D}^+\mathbf{U}^T = \mathbf{U}\,\text{diag}(\mathbf{I}, \mathbf{0})\mathbf{U}^T.$$

Their symmetry is obvious from their structure.

Solution of Problem 3

Note that the regression problem should be written as

$$y_i = \begin{bmatrix}1 & x_i\end{bmatrix}\begin{bmatrix}\vartheta_0\\ \vartheta_1\end{bmatrix}$$

and for all n samples of (x_i, y_i), we have the following definition :

$$\mathbf{y} = \begin{bmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{bmatrix}\begin{bmatrix}\vartheta_0\\ \vartheta_1\end{bmatrix}$$

See that firstly:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} = \frac{1}{n} \frac{1}{\overline{x^2} - \overline{x}^2} \begin{bmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{bmatrix}$$

On the other hand we have:

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} n\overline{y} \\ \sum x_i y_i \end{bmatrix}.$$

So finally the solution is given by:

$$\begin{aligned} (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} &= \frac{1}{n} \frac{1}{\overline{x^2} - \overline{x}^2} \begin{bmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{bmatrix} \begin{bmatrix} n\overline{y} \\ \sum x_i y_i \end{bmatrix} = \frac{1}{n} \frac{1}{\overline{x^2} - \overline{x}^2} \begin{bmatrix} n\overline{y}\overline{x^2} - \overline{x}(\sum x_i y_i) \\ -n\overline{y}.\overline{x} + \sum x_i y_i \end{bmatrix} \\ &= \frac{1}{\overline{x^2} - \overline{x}^2} \begin{bmatrix} \overline{y}\overline{x^2} - \overline{x}\rho_{xy} \\ -\overline{y}.\overline{x} + \rho_{xy} \end{bmatrix} = \frac{1}{\sigma_x^2} \begin{bmatrix} \overline{y}\overline{x^2} - \overline{x}\rho_{xy} \\ \sigma_{xy} \end{bmatrix} \end{aligned}$$

Therefore $\vartheta_1 = \frac{\sigma_{xy}}{\sigma_x^2}$ and

$$\vartheta_0 = \frac{1}{\sigma_x^2}(\overline{y}\overline{x^2} - \overline{x}\rho_{xy}) = \frac{1}{\sigma_x^2}(\overline{y}\overline{x^2} - \overline{x}(\overline{y}.\overline{x} + \sigma_{xy})) = \frac{1}{\sigma_x^2}(\overline{y}(\overline{x^2} - \overline{x}^2)) - \overline{x}\frac{\sigma_{xy}}{\sigma_x^2}$$

hence $\vartheta_0 = \overline{y} - \vartheta_1\overline{x}$.