Correction TD linear regression

Amaury Durand

November 28, 2018

Exercise 1 :

1. Let $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$, then $\forall \theta_0 \in \mathbb{R}$,

$$\left\| Y - X \begin{bmatrix} \theta_0 \\ \tilde{\boldsymbol{\theta}} \end{bmatrix} \right\| = \left\| Y - (1_n, \tilde{X}) \begin{bmatrix} \theta_0 \\ \tilde{\boldsymbol{\theta}} \end{bmatrix} \right\| = \left\| Y - \theta_0 1_n - \tilde{X}\tilde{\boldsymbol{\theta}} \right\| = \left\| Y_c - \tilde{X}_c \tilde{\boldsymbol{\theta}} + 1_n(\hat{\mu}_Y - \theta_0 - \hat{\mu}_X^\top \tilde{\boldsymbol{\theta}}) \right\|$$

Hence taking $\theta_0 = \hat{\mu}_Y - \hat{\mu}_X^\top \tilde{\boldsymbol{\theta}}$ we have

$$\left\| Y - X \begin{bmatrix} \hat{\mu}_Y - \hat{\mu}_X^\top \tilde{\boldsymbol{\theta}} \\ \tilde{\boldsymbol{\theta}} \end{bmatrix} \right\| = \left\| Y_c - \tilde{X}_c \tilde{\boldsymbol{\theta}} \right\|$$

Now, we have

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \|Y - X\boldsymbol{\theta}\| = \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \min_{\theta_0 \in \mathbb{R}} \left\| Y - X \begin{bmatrix} \theta_0 \\ \tilde{\boldsymbol{\theta}} \end{bmatrix} \right\| \leq \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \left\| Y - X \begin{bmatrix} \hat{\mu}_Y - \hat{\mu}_X^\top \tilde{\boldsymbol{\theta}} \\ \tilde{\boldsymbol{\theta}} \end{bmatrix} \right\| = \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \left\| Y_c - \tilde{X}_c \tilde{\boldsymbol{\theta}} \right\|$$

2. Let $Z = (z_1, \cdots, z_n)^\top \in \mathbb{R}^n$ and call $f : z \mapsto \|Z - z 1_n\|^2 = \sum_{i=1}^n (z_i - z)^2$, then f is convex and $\forall z \in \mathbb{R}$, $f'(z) = -2\sum_{i=1}^n (z_i - z)$ which is null if and only if $z = \frac{1}{n}\sum_{i=1}^n z_i = \overline{z}^n$.

3. Let $\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p$, then using question 2. with $Z = Y - \tilde{X}\tilde{\boldsymbol{\theta}}$ we get that $\forall \theta_0 \in \mathbb{R}$,

$$\left\| Y - \theta_0 1_n - \tilde{X}\tilde{\boldsymbol{\theta}} \right\| \geq \left\| Y - \hat{a}_n(\tilde{\boldsymbol{\theta}}) 1_n - \tilde{X}\tilde{\boldsymbol{\theta}} \right\|$$

with

$$\hat{a}_n(\tilde{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^n z_i = \frac{1}{n}\sum_{i=1}^n (y_i - x_i^\top \tilde{\boldsymbol{\theta}}) = \hat{\mu}_Y - \hat{\mu}_X^\top \tilde{\boldsymbol{\theta}}$$

4. Using the same calculus as in question 1.,

$$\begin{aligned}
\min_{\boldsymbol{\theta} \in \mathbb{R}^{p+1}} \|Y - X\boldsymbol{\theta}\| &= \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \min_{\theta_0 \in \mathbb{R}} \left\| Y - X \begin{bmatrix} \theta_0 \\ \tilde{\boldsymbol{\theta}} \end{bmatrix} \right\| \\
&= \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \min_{\theta_0 \in \mathbb{R}} \left\| Y - \theta_0 1_n - \tilde{X}\tilde{\boldsymbol{\theta}} \right\| \\
&\geq \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \left\| Y - \hat{a}_n(\tilde{\boldsymbol{\theta}}) 1_n - \tilde{X}\tilde{\boldsymbol{\theta}} \right\| \qquad \text{(by question 3.)} \\
&= \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \left\| Y - 1_n \hat{\mu}_Y + 1_n \hat{\mu}_X^\top \tilde{\boldsymbol{\theta}} - \tilde{X}\tilde{\boldsymbol{\theta}} \right\| \\
&= \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \left\| Y_c - \tilde{X}_c \tilde{\boldsymbol{\theta}} \right\|
\end{aligned}$$

An using question 1. we get the equality wanted.

5. The Hilbert projection theorem ensures that there is a unique $\hat{Y} = \operatorname{argmin}_{u \in \operatorname{Im}(X)} \|Y - u\|$ (it is the orthogonal projection of Y on the $\operatorname{Im}(X)$). By definition, we have $\hat{Y} = X\hat{\boldsymbol{\theta}}_n$. But we have shown in the previous question that

$$\left\|Y - \hat{Y}\right\| = \min_{\tilde{\boldsymbol{\theta}} \in \mathbb{R}^p} \left\|Y_c - \tilde{X}_c \tilde{\boldsymbol{\theta}}\right\| = \left\|Y - \hat{\mu}_Y 1_n - \tilde{X}_c \hat{\boldsymbol{\theta}}_{n,c}\right\|$$

Remarking that $\hat{\mu}_Y 1_n + \tilde{X}_c \hat{\boldsymbol{\theta}}_{n,c} = X(\hat{\mu}_Y - \hat{\mu}_X^T \hat{\boldsymbol{\theta}}_{n,c}, \hat{\boldsymbol{\theta}}_{n,c})$ belongs to $\operatorname{Im}(X)$ and using the uniqueness of the projection we simply have that

$$X\hat{\boldsymbol{\theta}}_n = X(\hat{\mu}_Y - \hat{\mu}_X^T \hat{\boldsymbol{\theta}}_{n,c}, \hat{\boldsymbol{\theta}}_{n,c})$$

which gives, using that $\ker(X) = \{0\}$, the final statement.

Exercise 2 :

1. We have

$$
\begin{aligned}
&(A+BCD)(A^{-1}-A^{-1}B(DA^{-1}B+C^{-1})^{-1}DA^{-1})\\
&= I_d + B(DA^{-1}B+C^{-1})^{-1}DA^{-1} + BCDA^{-1} - BCDA^{-1}B(DA^{-1}B+C^{-1})^{-1}DA^{-1}
\end{aligned}
$$

with

$$
\begin{aligned}
BCDA^{-1}B(DA^{-1}B+C^{-1})^{-1}DA^{-1} &= BC(DA^{-1}B+C^{-1}-C^{-1})(DA^{-1}B+C^{-1})^{-1}DA^{-1}\\
&= BCDA^{-1} + B(DA^{-1}B+C^{-1})^{-1}DA^{-1}
\end{aligned}
$$

Hence

$$
\begin{aligned}
&(A+BCD)(A^{-1}-A^{-1}B(DA^{-1}B+C^{-1})^{-1}DA^{-1})\\
&= I_d + B(DA^{-1}B+C^{-1})^{-1}DA^{-1} + BCDA^{-1} - BCDA^{-1} - B(DA^{-1}B+C^{-1})^{-1}DA^{-1}\\
&= I_d
\end{aligned}
$$

Which gives

$$(A+BCD)^{-1} = A^{-1}-A^{-1}B(DA^{-1}B+C^{-1})^{-1}DA^{-1}$$

2.

$$
\begin{aligned}
(X_{(n+1)}^\top X_{(n+1)})^{-1} &= \left([X_{(n)}^\top, \mathbf{x}_{n+1}] \begin{bmatrix} X_{(n)} \\ \mathbf{x}_{n+1}^\top \end{bmatrix}\right)^{-1}\\
&= \left(X_{(n)}^\top X_{(n)} + \mathbf{x}_{n+1}\mathbf{x}_{n+1}^\top\right)^{-1}\\
&\overset{(*)}{=} (X_{(n)}^\top X_{(n)})^{-1} - \frac{(X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}\mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}}{\mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}+1}\\
&\overset{(**)}{=} (X_{(n)}^\top X_{(n)})^{-1} - \frac{(X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}\left((X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}^\top\right)^\top}{\mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}+1}
\end{aligned}
$$

$(*)$: Using question 1. with $d=p+1,\ k=1,\ A=X_{(n)}^\top X_{(n)},\ B=\mathbf{x}_{n+1},\ C=I_1,\ D=\mathbf{x}_{n+1}^\top$
$(**)$: Because $(X_{(n)}^\top X_{(n)})^{-1}$ is symmetric since $X_{(n)}^\top X_{(n)}$ is.

We get the result wanted with $\zeta_{n+1} = (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}$ and $b_{n+1} = \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}$

3. $X_{(n+1)}^\top \mathbf{y}_{n+1} = [X_{(n)}^\top, x_{n+1}] \begin{bmatrix} \mathbf{y}_{(n)} \\ y_{n+1} \end{bmatrix} = X_{(n)}^\top \mathbf{y}_{(n)} + y_{n+1}\mathbf{x}_{n+1}$

4. Using the definition of $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_{n+1}$ and questions 2. and 3.

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}_{n+1} &= (X_{(n+1)}^\top X_{(n+1)})^{-1} X_{(n+1)}^\top \mathbf{y}_{(n+1)}\\
&= \left[(X_{(n)}^\top X_{(n)})^{-1} - \frac{\zeta_{n+1}\zeta_{n+1}^\top}{1+b_{n+1}}\right]\left[X_{(n)}^\top \mathbf{y}_{(n)} + y_{n+1}\mathbf{x}_{n+1}\right]\\
&= \underbrace{(X_{(n)}^\top X_{(n)})^{-1}X_{(n)}^\top \mathbf{y}_{(n)}}_{=\hat{\boldsymbol{\theta}}_n} + y_{n+1}\underbrace{(X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}}_{\zeta_{n+1}} - \frac{\zeta_{n+1}\zeta_{n+1}^\top X_{(n)}^\top \mathbf{y}_{(n)} + y_{n+1}\zeta_{n+1}\zeta_{n+1}^\top \mathbf{x}_{n+1}}{1+b_{n+1}}\\
&= \hat{\boldsymbol{\theta}}_n + \frac{1}{1+b_{n+1}}\left[\underbrace{(1+b_{n+1})y_{n+1}}_{\in\mathbb{R}}\zeta_{n+1} - \zeta_{n+1}\underbrace{\zeta_{n+1}^\top X_{(n)}^\top \mathbf{y}_{(n)}}_{\in\mathbb{R}} - y_{n+1}\zeta_{n+1}\underbrace{\zeta_{n+1}^\top \mathbf{x}_{n+1}}_{\in\mathbb{R}}\right]\\
&= \hat{\boldsymbol{\theta}}_n + \frac{1}{1+b_{n+1}}\left[(1+b_{n+1})y_{n+1} - \zeta_{n+1}^\top X_{(n)}^\top \mathbf{y}_{(n)} - y_{n+1}\zeta_{n+1}^\top \mathbf{x}_{n+1}\right]\zeta_{n+1}
\end{aligned}
$$

with

$$
\begin{aligned}
&(1+b_{n+1})y_{n+1} - \zeta_{n+1}^\top X_{(n)}^\top \mathbf{y}_{(n)} - y_{n+1}\zeta_{n+1}^\top \mathbf{x}_{n+1} \\
&= (1+\mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1})y_{n+1} - \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1} X_{(n)}^\top \mathbf{y}_{(n)} - y_{n+1}\mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1} \\
&= y_{n+1} - \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1} X_{(n)}^\top \mathbf{y}_{(n)} \\
&= y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_n
\end{aligned}
$$

Hence

$$\hat{\boldsymbol{\theta}}_{n+1} = \hat{\boldsymbol{\theta}}_n + \frac{u_{n+1}}{1+b_{n+1}}\zeta_{n+1}$$

with $u_{n+1} = y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_n$.

5. Using the result of question 4. we need to compute
   - $u_{n+1} = y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_n$ with one operation $(1, p+1) \times (p+1, 1)$
   - $b_{n+1} = \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}$ with one operation $(p+1, p+1) \times (p+1, 1)$ and one operation $(1, p+1) \times (p+1, 1)$
   - $\zeta_{n+1} = (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1}$ with one operation $(p+1, p+1) \times (p+1, 1)$

   We get a complexity of $O(p^2)$ instead of the $O(np^2)$ obtained when using the formula $\hat{\boldsymbol{\theta}}_{n+1} = (X_{(n+1)}^\top X_{(n+1)})^{-1} X_{(n+1)}^\top \mathbf{y}_{(n+1)}$

6. Multiplying the result of question 2. by $\mathbf{x}_{n+1}^\top$ on the left and $\mathbf{x}_{n+1}$ on the right we get

$$
\begin{aligned}
h_{n+1} := \mathbf{x}_{n+1}^\top (X_{(n+1)}^\top X_{(n+1)})^{-1}\mathbf{x}_{n+1} &= \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1} - \frac{\mathbf{x}_{n+1}^\top \zeta_{n+1}\zeta_{n+1}^\top \mathbf{x}_{n+1}}{1+b_{n+1}} \\
&= \mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1} - \frac{(\mathbf{x}_{n+1}^\top (X_{(n)}^\top X_{(n)})^{-1}\mathbf{x}_{n+1})^2}{1+b_{n+1}} \\
&= b_{n+1} - \frac{b_{n+1}^2}{1+b_{n+1}} \\
&= \frac{b_{n+1}}{1+b_{n+1}}
\end{aligned}
$$

   Hence $1 - h_{n+1} = \frac{1}{1+b_{n+1}}$ and finally $1 + b_{n+1} = \frac{1}{1-h_{n+1}}$

7. Using the result of question 6. and the given formula (which comes from multiplying by $\mathbf{x}_{n+1}^\top$ on the left the result of question 4.)

$$
\begin{aligned}
y_{n+1} - \hat{y}_{n+1} &= y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_n - \frac{u_{n+1}b_{n+1}}{1+b_{n+1}} \\
&= \underbrace{y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\theta}}_n}_{u_{n+1}} - u_{n+1}h_{n+1} \qquad \left(\text{because } h_{n+1} = \tfrac{b_{n+1}}{1+b_{n+1}}\right) \\
&= u_{n+1}(1 - h_{n+1})
\end{aligned}
$$

8. All the results above can be extended to the case where instead of removing the last row of X we remove the row at the i-th position. And we get by question 7.

$$y_i - \hat{y}_i = (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\theta}}_{(-i)})(1 - \hat{h}_i)$$

   hence

$$y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\theta}}_{(-i)} = \frac{y_i - \hat{y}_i}{1 - \hat{h}_i}$$

   and we get the result by replacing the expression in R_{cv}.