Exercices 6 solutions

1 OLS RISK DECOMPOSITION

Show the first part of proposition 15 in FTML.pdf (Risk decomposition for OLS, linear model, fixed design).

$$R_X(\theta) - R_X(\theta^*) = ||\theta - \theta^*||^2_{\hat{\Sigma}}$$

where $R_X(\theta)$ is the fixed design risk, defined by

$$R_X(\theta) = E_y\left[\frac{1}{n}||y - X\theta||^2\right] \tag{1}$$

with Σ the non-centered empirical covariance matrix :

$$\hat{\Sigma} = \frac{1}{n}X^T X \in \mathbb{R}^{d,d} \tag{2}$$

and the Mahalanobis distance norm.

$$||\theta||^2_{\hat{\Sigma}} = \theta^T \hat{\Sigma} \theta \tag{3}$$

1.0.1 *Solution*

We note that

$$\begin{aligned} R_X(\theta^*) &= E_y\left[\frac{1}{n}||y - X\theta^*||^2\right] \\ &= \frac{1}{n}E_\epsilon\left[||\epsilon||^2\right] \\ &= \frac{1}{n}E_\epsilon\left[\sum_{i=1}^{n}\epsilon_i^2\right] \\ &= \sigma^2 \end{aligned} \tag{4}$$

We now decompose $R_X(\theta)$:

$$\begin{aligned} R_X(\theta) &= E_y\left[\frac{1}{n}||y - X\theta||^2\right] \\ &= \frac{1}{n}E_y\left[||y - X\theta^* + X\theta^* - X\theta||^2\right] \end{aligned} \tag{5}$$

For for any vectors z and $z' \in \mathbb{R}^n$, we have

$$\begin{aligned} ||z + z'||^2 &= \langle z + z', z + z' \rangle \\ &= \langle z, z \rangle + 2\langle z, z' \rangle + \langle z', z' \rangle \\ &= ||z||^2 + 2\langle z, z' \rangle + ||z'||^2 \end{aligned} \tag{6}$$

Hence,

$$
\begin{aligned}
R_X(\theta) &= E_y\left[\frac{1}{n}||y - X\theta^*||^2 + \frac{2}{n}\langle y - X\theta^*, X\theta^* - X\theta\rangle + \frac{1}{n}||X\theta^* - X\theta||^2\right] \\
&= R_X(\theta^*) + \frac{2}{n}E_y\left[\langle y - X\theta^*, X\theta^* - X\theta\rangle\right] + E_y\left[\frac{1}{n}||X(\theta^* - \theta)||^2\right] \\
&= R_X(\theta^*) + \frac{2}{n}E_y\left[\langle \epsilon, X(\theta^* - \theta)\rangle\right] + \frac{1}{n}||X(\theta^* - \theta)||^2
\end{aligned} \tag{7}
$$

But using that for all $z \in \mathbb{R}^n$, $||z||^2 = \langle z, z\rangle = z^T z$, the last term writes :

$$
\begin{aligned}
\frac{1}{n}||X(\theta^* - \theta)||^2 &= \frac{1}{n}\left(X(\theta^* - \theta)\right)^T\left(X(\theta^* - \theta)\right) \\
&= \left(\theta^* - \theta\right)^T \frac{1}{n}X^T X\left(\theta^* - \theta\right) \\
&= ||\theta - \theta^*||^2_{\hat{\Sigma}}
\end{aligned} \tag{8}
$$

We now focus on the second term of the sum :

$$
\begin{aligned}
E_y\left[\langle \epsilon, X(\theta^* - \theta)\rangle\right] &= E_y\left[\epsilon^T, X(\theta^* - \theta)\right] \\
&= (E_y\left[\epsilon\right])^T X(\theta^* - \theta) \\
&= 0
\end{aligned} \tag{9}
$$

This concludes the proof.