Tarea n°4

Consideremos el modelo de regresión

$$Y = X\beta + \epsilon$$

donde supongamos que X es una matriz de $\mathcal{M}_{n,k}(\mathbb{R})$ determinista, y que $\mathbb{E}[\epsilon] = 0$, $\mathbb{E}\left[\epsilon\epsilon^T\right] = \sigma^2 I_n$. El vector de parámetros $\beta \in \mathbb{R}^k$ es de dimensión k.

1. Supongamos que $k > n$. Que decir del estimador de mínimos cuadrados en este caso ? Que decir de la matriz $X^T X$?
2. Estamos interesados en el estimador que para todo $\lambda > 0$ está definido por

$$\hat{\beta}_\lambda = \mathrm{argmin}_{\beta \in \mathbb{R}^k}(\|Y - X\beta\|^2 + \lambda\|\beta^2\|)$$

   Mostrar que la solución es única y que es también solución de la ecuación lineal $-2X^T(Y - X\beta) + 2\lambda\beta = 0$.
3. En este inciso queremos dar una formula explicita para la solución.
   (a) Preliminar : Mostrar que para cada $\lambda > 0$, la matriz $X^T X + \lambda I_n$ es invertible.
   (b) Concluir dando la forma explicita de la solución.
4. Calcular el sesgo y la matriz de covarianza de la solución $\hat{\beta}_\lambda$.
5. Finalmente supongamos $k = 1$ de tal manera que el problema se reduce a la regresión sencilla. Muestre que hay un valor de λ tal que el riesgo del estimador de parámetro λ es menor que el riesgo del estimador de mínimos cuadrados.