High-dimensional HW#4

102225014 Ai-Chun Chen

Q. Find the best value of "a" that minimize

$$\mathrm{MSE}\{\hat{\boldsymbol\beta}(\mathrm{a})\} = \mathrm{E}\{\hat{\boldsymbol\beta}(\mathrm{a}) - \boldsymbol\beta\}^{\mathrm{T}}\{\hat{\boldsymbol\beta}(\mathrm{a}) - \boldsymbol\beta\}$$

under $\mathbf{y} = \mathbf{X}\boldsymbol\beta + \boldsymbol\varepsilon$, $\mathbf{X} = [\,\mathbf{x}_1 \cdots \mathbf{x}_p\,]$ (without intercept), $\boldsymbol\beta = (\beta_1 \cdots \beta_p)^{\mathrm{T}}$,

$\boldsymbol\varepsilon \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$ and $\hat{\boldsymbol\beta}(\mathrm{a}) = \mathrm{a}\,\hat{\boldsymbol\beta}^{\mathrm{OLS}}$.

<u>Solution</u>:

$$\begin{aligned}
&\mathrm{MSE}\{\hat{\boldsymbol\beta}(\mathrm{a})\}\\
&= \mathrm{tr}[\,\mathrm{Var}\{\hat{\boldsymbol\beta}(\mathrm{a})\}\,] + \left\|\mathrm{bias}\{\hat{\boldsymbol\beta}(\mathrm{a})\}\right\|^2\\
&= \mathrm{tr}[\,\mathrm{Var}\{a(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}\}\,] + \left\|\mathrm{E}\{a(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} - \boldsymbol\beta\}\right\|^2\\
&= a^2\mathrm{tr}\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathrm{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\} + \|a\boldsymbol\beta - \boldsymbol\beta\|^2\\
&= a^2\sigma^2\mathrm{tr}\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\} + (\mathrm{a}-1)^2\|\boldsymbol\beta\|^2\\
&= a^2\sigma^2\mathrm{tr}(\boldsymbol\Lambda^{-1}) + (\mathrm{a}-1)^2\|\boldsymbol\beta\|^2 \qquad \text{where } \boldsymbol\Lambda \text{ is the diagonal matrix with all diagonal}\\
&= a^2\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} + (\mathrm{a}-1)^2\|\boldsymbol\beta\|^2 \qquad \text{elements be the eigenvalues of } \mathbf{X}^{\mathrm{T}}\mathbf{X}, \lambda_1, \cdots, \lambda_p.
\end{aligned}$$

Taking derivative with respect to a to find the extreme value of MSE.

$$\Rightarrow \frac{d}{da}\mathrm{MSE}\{\hat{\boldsymbol\beta}(a)\} = 2a\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} + 2(\mathrm{a}-1)\|\boldsymbol\beta\|^2 \overset{let}{=} 0$$

$$\Rightarrow a\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} + (\mathrm{a}-1)\|\boldsymbol\beta\|^2 = 0$$

$$\Rightarrow a\left(\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} + \|\boldsymbol\beta\|^2\right) - \|\boldsymbol\beta\|^2 = 0$$

$$\Rightarrow a = \frac{\|\boldsymbol\beta\|^2}{\left(\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} + \|\boldsymbol\beta\|^2\right)}$$

Check the 2nd derivative,

$$\Rightarrow \frac{d^2}{d^2a}\mathrm{MSE}\{\hat{\boldsymbol\beta}(a)\} = 2\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} + 2\|\boldsymbol\beta\|^2 > 0$$

$$\Rightarrow \mathrm{MSE}\{\hat{\boldsymbol\beta}(a)\} \text{ is minimized when } a = \frac{\|\boldsymbol\beta\|^2}{\left(\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i} + \|\boldsymbol\beta\|^2\right)}.$$