Tarea 2

Temas selectos de estadística
Aprendizaje de máquina probabilístico
Semestre 2025-1

Diego Leipen Lara

1. En el modelo de regresión lineal

$$y = X_0\beta_0 + \epsilon_0, \qquad \epsilon_0 \sim \mathcal{N}_n(0, \sigma_0^2 I_n) \tag{1}$$

donde X tiene dimensión $n \times p$, al estar considerando intercepto, y β_0 es tal que, con $0 < p_0 \leq p$, sus últimas $p - p_0$ entradas son cero. Más aún, utilizamos el modelo

$$y = X\beta + \epsilon, \qquad \epsilon \sim \mathcal{N}_n(0, \sigma^2 I_n). \tag{2}$$

Sean $\theta_0 = (\beta_0', \sigma_0^2)$ y $\theta = (\beta', \sigma^2)$. El criterio de información de Akaike corregido para el modelo (2) es

$$\mathrm{AIC}_c \coloneqq -2\log\left(f(y;\theta)\right) + \frac{2n\dim\theta}{n - \dim\theta - 1}.$$

1. Muestre que en general

$$\mathrm{AIC}_c = \mathrm{AIC} + \frac{2\dim\theta^2 + 2\dim\theta}{n - \dim\theta - 1}.$$

2. Para el modelo (2) utilizando los estimadores

$$\hat\beta_n \coloneqq (X'X)^{-1}Xy, \qquad \hat\sigma_n^2 \coloneqq \frac{1}{n}(y - X\hat\beta_n)'(y - X\hat\beta_n)$$

exprese analíticamente el AIC_c.

3. Muestre que

$$\mathbb{E}_0\left[-2\log\left(f(y;\theta_0)\right)\right] - \mathbb{E}_0\left[-2\log\left(f(y;\hat\theta_n)\right)\right] = -\mathbb{E}_0\left[n\log\left(\frac{\hat\sigma_n^2}{\sigma_0^2}\right)\right].$$

4. Muestre que

$$\mathbb{E}_0\left[\mathbb{E}_0\left[-2\log\left(f(y;\theta)\right)\right]|_{\theta=\hat\theta_n}\right] - \mathbb{E}_0\left[-2\log\left(f(y;\theta_0)\right)\right] = \mathbb{E}_0\left[n\log\left(\frac{\hat\sigma_n^2}{\sigma_0^2}\right)\right] + \frac{2(p+1)n}{n-p-2}.$$

5. Para la discrepancia de Kullback

$$d\left(\hat\theta_n, \theta_0\right) = \mathbb{E}_0\left[-2\log\left(f(y;\theta)\right)\right]|_{\theta=\hat\theta_n}$$

muestre que

$$\mathbb{E}_0[\mathrm{AIC}_c] = \mathbb{E}_0\left[d\left(\hat\theta_n, \theta_0\right)\right].$$

Demostración.

1. Calculando directamente,

$$\begin{aligned}
&\mathrm{AIC} + \frac{2\dim\theta^2 + 2\dim\theta}{n - \dim\theta - 1} \\
&= -2\log\left(f(y;\hat{\theta}_n)\right) + 2\dim\theta + \frac{2\dim\theta^2 + 2\dim\theta}{n-\dim\theta-1} \\
&= -2\log\left(f(y;\hat{\theta}_n)\right) + \frac{2\dim\theta(n-\dim\theta-1) + 2\dim\theta^2 + 2\dim\theta}{n-\dim\theta-1} \\
&= -2\log\left(f(y;\hat{\theta}_n)\right) + \frac{2\dim\theta n}{n-\dim\theta-1} \\
&= \mathrm{AIC}_c\,.
\end{aligned}$$

2. Calculando directamente,

$$f(y;\hat{\theta}_n) = \frac{1}{(2\pi\hat{\sigma}_n^2)^{n/2}}\exp\left\{-\frac{1}{2\hat{\sigma}_n^2}(y - X\hat{\beta}_n)'(y - X\hat{\beta}_n)\right\} = \frac{1}{(2\pi\hat{\sigma}_n^2)^{n/2}}\exp\left\{-\frac{n}{2}\right\}.$$

Entonces

$$\log\left(f(y;\hat{\theta}_n)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}_n^2) - \frac{n}{2}. \tag{3}$$

Por lo tanto,

$$\begin{aligned}
\mathrm{AIC}_c &= -2\log\left(f(y;\theta)\right) + \frac{2n\dim\theta}{n-\dim\theta-1} \\
&= -n\log(2\pi) - n\log(\hat{\sigma}_n^2) - n + \frac{2n\dim\theta}{n-\dim\theta-1}.
\end{aligned}$$

3. Primero notemos que por (3)

$$\mathbb{E}_0\left[-2\log\left(f(y;\hat{\theta}_n)\right)\right] = n\log(2\pi) + \mathbb{E}_0\left[n\log(\hat{\sigma}_n^2)\right] + n. \tag{4}$$

Por otro lado,

$$f(y;\theta_0) = \frac{1}{(2\pi\hat{\sigma}_n^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma_0^2}(y - X_0\beta_0)'(y - X_0\beta_0)\right\}.$$

Entonces

$$\log\left(f(y,\theta_0)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_0^2) - \frac{1}{2\sigma_0^2}(y - X_0\beta_0)'(y - X_0\beta_0).$$

Más aun,

$$\begin{aligned}
\mathbb{E}_0\left[(y - X_0\beta_0)'(y - X_0\beta_0)\right] &= \mathbb{E}_0\left[\sum_{i=1}^n (y_i - (X_0\beta_0)_i)^2\right] \\
&= \sum_{i=1}^n \mathbb{E}_0\left[(y_i - (X_0\beta_0)_i)^2\right] \\
&= \sum_{i=1}^n \left\{\mathrm{Var}_0\left[y_i - (X_0\beta_0)_i\right] + \left(\mathbb{E}_0\left[y_i - (X_0\beta_0)_i\right]\right)^2\right\} \\
&= \sum_{i=1}^n \left\{\sigma_0^2 + 0^2\right\} \\
&= n\sigma_0^2
\end{aligned}$$

donde la cuarta igualdad se cumple porque $\mathrm{Var}_0\left[y_i - (X_0\beta_0)_i\right] = \mathrm{Var}_0\left[y_i\right] = \sigma_0^2$. Entonces

$$\mathbb{E}_0\left[-2\log\left(f(y;\theta_0)\right)\right] = n\log(2\pi) + \mathbb{E}_0\left[n\log(\sigma_0^2)\right] + n. \tag{5}$$

Juntando (4) y (5) obtenemos

$$\begin{aligned}\mathbb{E}_0\left[-2\log\left(f(y;\theta_0)\right)\right] - \mathbb{E}_0\left[-2\log\left(f(y;\hat{\theta}_n)\right)\right] &= \mathbb{E}_0\left[n\log(\sigma_0^2)\right] - \mathbb{E}_0\left[n\log(\hat{\sigma}_n^2)\right] \\ &= -\mathbb{E}_0\left[n\log\left(\frac{\hat{\sigma}_n^2}{\sigma_0^2}\right)\right].\end{aligned}$$

4. Calculando directamente,

$$f(y;\theta) = \frac{1}{(2\pi\hat{\sigma}_n^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}.$$

Entonces

$$\log\left(f(y,\theta)\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

Más aun,

$$\begin{aligned}\mathbb{E}_0\left[(y - X\beta)'(y - X\beta)\right] &= \mathbb{E}_0\left[\sum_{i=1}^n (y_i - (X\beta)_i)^2\right] \\ &= \sum_{i=1}^n \mathbb{E}_0\left[(y_i - (X\beta)_i)^2\right] \\ &= \sum_{i=1}^n \left\{\mathrm{Var}_0\left[y_i - (X\beta)_i\right] + \left(\mathbb{E}_0\left[y_i - (X\beta)_i\right]\right)^2\right\} \\ &= \sum_{i=1}^n \left\{\sigma_0^2 + \left((X_0\beta_0)_i - (X\beta)_i\right)^2\right\} \\ &= n\sigma_0^2 + (X_0\beta_0 - X\beta)'(X_0\beta_0 - X\beta).\end{aligned}$$

Entonces

$$\begin{aligned}\mathbb{E}_0\left[-2\log\left(f(y,\theta)\right)\right] &= n\log(2\pi) + n\log(\sigma^2) + \frac{1}{\sigma^2}\left(n\sigma_0^2 + (X_0\beta_0 - X\beta)'(X_0\beta_0 - X\beta)\right) \\ &= n\log(2\pi) + n\log(\sigma^2) + \frac{n\sigma_0^2}{\sigma^2} + \frac{(X_0\beta_0 - X\beta)'(X_0\beta_0 - X\beta)}{\sigma^2}.\end{aligned}$$

Entonces

$$\mathbb{E}_0\left[-2\log\left(f(y,\theta)\right)\right]|_{\theta=\hat{\theta}_n} = n\log(2\pi) + n\log(\hat{\sigma}_n^2) + \frac{n\sigma_0^2}{\hat{\sigma}_n^2} + \frac{(X_0\beta_0 - X\hat{\beta}_n)'(X_0\beta_0 - X\hat{\beta}_n)}{\hat{\sigma}_n^2}.$$

Pero en el articulo *Estimation Optimality of Corrected AIC and Modified Cp in Linear Regression* de Davies, Neath, y Cavanaugh (justo despues de la ecuación 3.3) se menciona que

$$\frac{n\sigma_0^2}{\hat{\sigma}_n^2} \sim \chi_{n-p}^2,$$

$$\frac{(X_0\beta_0 - X\hat{\beta}_n)'(X_0\beta_0 - X\hat{\beta}_n)}{\sigma_0^2} \sim \chi_p^2.$$

Entonces

$$\mathbb{E}_0\left[\frac{n\sigma_0^2}{\hat{\sigma}_n^2}\right] = n^2\mathbb{E}_0\left[\frac{\sigma_0^2}{n\hat{\sigma}_n^2}\right] = n^2\left(\frac{1}{n-p-2}\right)$$

donde la ultima igualdad se cumple porque si $Z \sim \chi_k^2$, entonces $\mathbb{E}[\frac{1}{Z}] = \frac{1}{k-2}$. Más aun,

$$\begin{aligned}
\mathbb{E}_0\left[\frac{(X_0\beta_0 - X\hat{\beta}_n)'(X_0\beta_0 - X\hat{\beta}_n)}{\hat{\sigma}_n^2}\right] &= n\mathbb{E}_0\left[\frac{\sigma_0^2}{n\hat{\sigma}_n^2}\frac{(X_0\beta_0 - X\hat{\beta}_n)'(X_0\beta_0 - X\hat{\beta}_n)}{\sigma_0^2}\right] \\
&= n\mathbb{E}_0\left[\frac{\sigma_0^2}{n\hat{\sigma}_n^2}\right]\mathbb{E}_0\left[\frac{(X_0\beta_0 - X\hat{\beta}_n)'(X_0\beta_0 - X\hat{\beta}_n)}{\sigma_0^2}\right] \\
&= n\left(\frac{1}{n-p-2}\right)p
\end{aligned}$$

donde la segunda igualdad se cumple por independencia de $\hat{\beta}_n$ y $\hat{\sigma}_n^2$. De esto se sigue que

$$\begin{aligned}
&\mathbb{E}_0\left[\mathbb{E}_0\left[-2\log\left(f(y,\theta)\right)\right]|_{\theta=\hat{\theta}_n}\right] \\
&\quad= \mathbb{E}_0\left[n\log(2\pi) + n\log(\hat{\sigma}_n^2) + \frac{n\sigma_0^2}{\hat{\sigma}_n^2} + \frac{(X_0\beta_0 - X\hat{\beta}_n)'(X_0\beta_0 - X\hat{\beta}_n)}{\hat{\sigma}_n^2}\right] \\
&\quad= n\log(2\pi) + \mathbb{E}_0\left[n\log(\hat{\sigma}_n^2)\right] + \frac{n^2}{n-p-2} + \frac{np}{n-p-2}.
\end{aligned}$$

Entonces

$$\begin{aligned}
&\mathbb{E}_0\left[\mathbb{E}_0\left[-2\log\left(f(y,\theta)\right)\right]|_{\theta=\hat{\theta}_n}\right] - \mathbb{E}_0\left[-2\log\left(f(y;\theta_0)\right)\right] \\
&= n\log(2\pi) + \mathbb{E}_0\left[n\log(\hat{\sigma}_n^2)\right] + \frac{n^2+np}{n-p-2} - \left(n\log(2\pi) + \mathbb{E}_0\left[n\log(\sigma_0^2)\right] + n\right) \qquad \text{(cf. (5))} \\
&= \mathbb{E}_0\left[n\log\left(\frac{\hat{\sigma}_n^2}{\sigma_0^2}\right)\right] + \frac{2(p+1)n}{n-p-2}.
\end{aligned}$$

5. Calculando directamente,

$$\begin{aligned}
\mathbb{E}_0\left[d\left(\hat{\theta}_n,\theta_0\right)\right] &= \mathbb{E}_0\left[\mathbb{E}_0\left[-2\log\left(f(y;\theta)\right)\right]|_{\theta=\hat{\theta}_n}\right] \\
&= \mathbb{E}_0\left[-2\log\left(f(y;\theta_0)\right)\right] + \mathbb{E}_0\left[n\log\left(\frac{\hat{\sigma}_n^2}{\sigma_0^2}\right)\right] + \frac{2(p+1)n}{n-p-2} \qquad \text{(por el inciso 4)} \\
&= \mathbb{E}_0\left[-2\log\left(f(y;\hat{\theta}_n)\right)\right] + \frac{2(p+1)n}{n-p-2} \qquad \text{(por el inciso 3)} \\
&= \mathbb{E}_0[\text{AIC}_c].
\end{aligned}$$

□

2. Demuestre que para el problema de optimización

$$\min_{\beta\in\mathbb{R}^p}\left\{\frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\beta\|_1\right\}$$

en caso de que X sea ortonormal, entonces

$$\beta^{\text{LASSO}}=\operatorname{sign}(\hat\beta_i)\left(|\hat\beta_i|-\lambda\right)^+$$

con $\hat\beta=(X'X)^{-1}X'y$.

Demostración. Como X es ortonormal, $\hat\beta=(X'X)^{-1}X'y=X'y$. Entonces

$$\begin{aligned}
\|y-X\beta\|_2^2&=(y-X\beta)'(y-X\beta)\\
&=(y'-\beta'X')(y-X\beta)\\
&=y'y-y'X\beta-\beta'X'y+\beta'X'X\beta\\
&=\hat\beta'\hat\beta-\hat\beta'\beta-\beta'\hat\beta+\beta'\beta\\
&=\|\hat\beta\|_2^2-2\sum_{i=1}^p\hat\beta_i\beta_i+\sum_{i=1}^p\beta_i^2.
\end{aligned}$$

Entonces

$$\begin{aligned}
&\min_{\beta\in\mathbb{R}^p}\left\{\frac{1}{2}\|y-X\beta\|_2^2+\lambda\|\beta\|_1\right\}\\
&=\min_{\beta\in\mathbb{R}^p}\left\{\frac{1}{2}\left(\|\hat\beta\|_2^2-2\sum_{i=1}^p\hat\beta_i\beta_i+\sum_{i=1}^p\beta_i^2\right)+\lambda\|\beta\|_1\right\}\\
&=\min_{\beta\in\mathbb{R}^p}\left\{\frac{1}{2}\left(-2\sum_{i=1}^p\hat\beta_i\beta_i+\sum_{i=1}^p\beta_i^2\right)+\lambda\|\beta\|_1\right\}\\
&=\min_{\beta\in\mathbb{R}^p}\left\{\sum_{i=1}^p\left(-2\hat\beta_i\beta_i+\beta_i^2+2\lambda|\beta_i|\right)\right\}\\
&=\sum_{i=1}^p\min_{\beta_i\in\mathbb{R}}\left\{-2\hat\beta_i\beta_i+\beta_i^2+2\lambda|\beta_i|\right\}.
\end{aligned}$$

donde la segunda igualdad se cumple porque $\|\hat\beta\|_2^2$ no depende de β. Sean

$$\begin{aligned}
f_i(\beta_i)&=-2\hat\beta_i\beta_i+\beta_i^2+2\lambda\beta_i,\\
g_i(\beta_i)&=-2\hat\beta_i\beta_i+\beta_i^2-2\lambda\beta_i.
\end{aligned}$$

Entonces

$$\min_{\beta_i\in\mathbb{R}}\left\{-2\hat\beta_i\beta_i+\beta_i^2+2\lambda|\beta_i|\right\}=\min\left\{\min_{\beta_i\geq0}f_i(\beta_i),\min_{\beta_i\leq0}g_i(\beta_i)\right\}.$$

Además,

$$\begin{aligned}
f_i'(\beta_i)&=-2\hat\beta_i+2\beta_i+2\lambda,\\
g_i'(\beta_i)&=-2\hat\beta_i+2\beta_i-2\lambda.
\end{aligned}$$

Observación 1. Si $\hat{\beta}_i \leq \lambda$, entonces para todo $\beta_i \geq 0$

$$f_i'(\beta_i) = -2\hat{\beta}_i + 2\beta_i + 2\lambda \geq -2\lambda + 2\beta_i + 2\lambda = 2\beta_i \geq 0.$$

En particular, f_i es creciente en $(0, \infty)$ y $\min_{\beta_i \geq 0} f_i(\beta_i) = 0$.

Observación 2. Si $\hat{\beta}_i \geq -\lambda$, entonces para todo $\beta_i \leq 0$

$$g_i'(\beta_i) = -2\hat{\beta}_i + 2\beta_i - 2\lambda \leq 2\lambda + 2\beta_i - 2\lambda = 2\beta_i \leq 0.$$

En particular, g_i es decreciente en $(-\infty, 0)$ y $\min_{\beta_i \leq 0} g_i(\beta_i) = 0$.

Caso 1. $\hat{\beta}_i \in [-\lambda, \lambda]$.
Entonces por las observaciones anteriores

$$\min_{\beta_i \in \mathbb{R}} \left\{ -2\hat{\beta}_i\beta_i + \beta_i^2 + 2\lambda|\beta_i| \right\} = \min \left\{ \min_{\beta_i \geq 0} f_i(\beta_i), \min_{\beta_i \leq 0} g_i(\beta_i) \right\} = \min\{0, 0\} = 0.$$

Caso 2. $\hat{\beta}_i \in (-\infty, -\lambda)$.
Entonces $\hat{\beta}_i + \lambda < 0$ y es fácil verificar que $g_i'(\hat{\beta}_i + \lambda) = 0$. Usando esto y el hecho de que g_i es una función cuadrática convexa (el coeficiente de β_i^2 es $+1$) obtenemos la primera de las siguientes igualdades.

$$\begin{aligned}
\min_{\beta_i \leq 0} g_i(\beta_i) &= g_i(\hat{\beta}_i + \lambda) \\
&= -2\hat{\beta}_i(\hat{\beta}_i + \lambda) + (\hat{\beta}_i + \lambda)^2 - 2\lambda(\hat{\beta}_i + \lambda) \\
&= -2\hat{\beta}_i^2 - 2\hat{\beta}_i\lambda + \hat{\beta}_i^2 + 2\hat{\beta}_i\lambda + \lambda^2 - 2\lambda\hat{\beta}_i - 2\lambda^2 \\
&= -\hat{\beta}_i^2 - 2\hat{\beta}_i\lambda - \lambda^2 \\
&= -(\hat{\beta}_i + \lambda)^2.
\end{aligned}$$

Por otro lado, como $\hat{\beta}_i \in (-\infty, -\lambda)$, en particular $\hat{\beta}_i \leq \lambda$ y por la Obs. 1, $\min_{\beta_i \geq 0} f_i(\beta_i) = 0$. Por lo tanto,

$$\min_{\beta_i \in \mathbb{R}} \left\{ -2\hat{\beta}_i\beta_i + \beta_i^2 + 2\lambda|\beta_i| \right\} = \min \left\{ \min_{\beta_i \geq 0} f_i(\beta_i), \min_{\beta_i \leq 0} g_i(\beta_i) \right\} = \min\{0, -(\hat{\beta}_i + \lambda)^2\} = -(\hat{\beta}_i + \lambda)^2.$$

Caso 3. $\hat{\beta}_i \in (\lambda, \infty)$.
Entonces $\hat{\beta}_i - \lambda > 0$ y es fácil verificar que $f_i'(\hat{\beta}_i - \lambda) = 0$. Usando esto y el hecho de que f_i es una función cuadrática convexa (el coeficiente de β_i^2 es $+1$) obtenemos la primera de las siguientes igualdades.

$$\begin{aligned}
\min_{\beta_i \leq 0} f_i(\beta_i) &= f_i(\hat{\beta}_i - \lambda) \\
&= -2\hat{\beta}_i(\hat{\beta}_i - \lambda) + (\hat{\beta}_i - \lambda)^2 + 2\lambda(\hat{\beta}_i - \lambda) \\
&= -2\hat{\beta}_i^2 + 2\hat{\beta}_i\lambda + \hat{\beta}_i^2 - 2\hat{\beta}_i\lambda + \lambda^2 + 2\lambda\hat{\beta}_i - 2\lambda^2 \\
&= -\hat{\beta}_i^2 + 2\hat{\beta}_i\lambda - \lambda^2 \\
&= -(\hat{\beta}_i - \lambda)^2.
\end{aligned}$$

Por otro lado, como $\hat{\beta}_i \in (\lambda, \infty)$, en particular $\hat{\beta}_i \geq -\lambda$ y por la Obs. 2, $\min_{\beta_i \geq 0} g_i(\beta_i) = 0$. Por lo tanto,

$$\min_{\beta_i \in \mathbb{R}} \left\{ -2\hat{\beta}_i\beta_i + \beta_i^2 + 2\lambda|\beta_i| \right\} = \min \left\{ \min_{\beta_i \geq 0} f_i(\beta_i), \min_{\beta_i \leq 0} g_i(\beta_i) \right\} = \min\{-(\hat{\beta}_i - \lambda)^2, 0\} = -(\hat{\beta}_i - \lambda)^2.$$

En resumen,

$$\min_{\beta_i \in \mathbb{R}} \left\{ -2\hat{\beta}_i \beta_i + \beta_i^2 + 2\lambda |\beta_i| \right\} = \begin{cases} 0 & \text{si } \hat{\beta}_i \in [-\lambda, \lambda] \\ -(\hat{\beta}_i + \lambda)^2 & \text{si } \hat{\beta}_i \in (-\infty, -\lambda) \\ -(\hat{\beta}_i - \lambda)^2 & \text{si } \hat{\beta}_i \in (\lambda, \infty) \end{cases}$$

Por lo tanto,

$$\begin{aligned} \beta_i^{\text{LASSO}} &= \begin{cases} 0 & \text{si } \hat{\beta}_i \in [-\lambda, \lambda] \\ \hat{\beta}_i + \lambda & \text{si } \hat{\beta}_i \in (-\infty, -\lambda) \\ \hat{\beta}_i - \lambda & \text{si } \hat{\beta}_i \in (\lambda, \infty) \end{cases} \\ &= \operatorname{sign}(\hat{\beta}_i) \left(|\hat{\beta}_i| - \lambda \right)^+ . \end{aligned}$$

$\square$

Lema 1. Sea $f : \mathbb{R}^p \to \mathbb{R}$ tal que $\arg\min_{\beta \in \mathbb{R}^p} f(\beta)$ es único. Entonces para cualquier $c \in \mathbb{R}$,

$$\arg\min_{\beta \in \mathbb{R}^p} f(c\beta) = \frac{1}{c} \arg\min_{\beta \in \mathbb{R}^p} f(\beta).$$

Demostración. Denotemos $\beta_0 = \arg\min_{\beta \in \mathbb{R}^p} f(c\beta)$. Por definición,

$$f(c\beta_0) \leq f(c\beta) \qquad \forall \beta \in \mathbb{R}^p.$$

En particular,

$$f(c\beta_0) \leq f(\beta) \qquad \forall \beta \in \mathbb{R}^p.$$

Por lo tanto, $c\beta_0 = \arg\min_{\beta \in \mathbb{R}^p} f(\beta)$. $\square$

3. La penalización de elastic net esta dada por el problema de optimización

$$b_1 = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \|y - X\beta\|_2^2 + \lambda_2 \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 \right\}.$$

Consideremos alternativamente el problema de optimización

$$b_2 = \arg\min_{\beta \in \mathbb{R}^p} \left\{ \|\tilde{y} - \tilde{X}\beta\|_2^2 + c\lambda_1 \|\beta\|_1 \right\}$$

donde

$$c = (1 + \lambda_2)^{-1/2}, \qquad \tilde{X} = c \begin{pmatrix} X \\ \sqrt{\lambda_2} I_p \end{pmatrix}, \qquad \tilde{y} = \begin{pmatrix} y \\ 0_{p \times 1} \end{pmatrix}.$$

Muestre que $b_1 = cb_2$ y explique cómo se puede entonces resolver elastic net mediante la solución de Lasso.

Demostración. Calculando directamente,

$$\begin{aligned}
\|\tilde{y} - \tilde{X}\beta\|_2^2 &= \sum_{i=1}^{n+p} (\tilde{y}_i - (\tilde{X}\beta)_i)^2 \\
&= \sum_{i=1}^{n} (\tilde{y}_i - (\tilde{X}\beta)_i)^2 + \sum_{i=n+1}^{n+p} (\tilde{y}_i - (\tilde{X}\beta)_i)^2 \\
&= \sum_{i=1}^{n} (y_i - c(X\beta)_i)^2 + \sum_{i=n+1}^{n+p} (0 - c\sqrt{\lambda_2}\beta_{i-n})^2 \\
&= \|y - X(c\beta)\|_2^2 + \lambda_2 \|c\beta\|_2^2.
\end{aligned}$$

Además, como $c > 0$, entonces $c\lambda_1 \|\beta\|_1 = \lambda_1 \|c\beta\|_1$. Por lo tanto,

$$\begin{aligned}
b_2 &= \arg\min_{\beta \in \mathbb{R}^p} \left\{ \|\tilde{y} - \tilde{X}\beta\|_2^2 + c\lambda_1 \|\beta\|_1 \right\} \\
&= \arg\min_{\beta \in \mathbb{R}^p} \left\{ \|y - X(c\beta)\|_2^2 + \lambda_2 \|c\beta\|_2^2 + \lambda_1 \|c\beta\|_1 \right\} \\
&= \frac{1}{c} b_1
\end{aligned}$$

donde la última igualdad es por el lema 1. Finalmente, podemos resolver elastic net mediante el problema alternativo, el cual es un problema de tipo Lasso. $\square$