Tarea 1

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

Diego Leipen Lara

1. Demuestre que en el modelo Bayesiano

$$\begin{aligned} \mu_i &\overset{\text{iid}}{\sim} \mathcal{N}(m, \sigma^2), \qquad 1 \le i \le n, \\ x_i|\mu_i &\overset{\text{iid}}{\sim} \mathcal{N}(\mu_i, 1), \qquad 1 \le i \le n. \end{aligned}$$

se tiene que

$$\mathbb{E}\left\{\|\hat{\mu}^{(JS)} - \mu\|^2\right\} = \frac{n\sigma^2}{\sigma^2+1} + \frac{3}{\sigma^2+1}$$

con

$$\hat{\mu}^{(JS)} = \overline{x}\mathbf{1} + \left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x - \overline{x}\mathbf{1}).$$

Demostración. Sabemos que

$$\begin{aligned} x_i &\overset{\text{iid}}{\sim} \mathcal{N}(m, \sigma^2+1), \qquad 1 \le i \le n, \\ \mu_i|x_i &\overset{\text{iid}}{\sim} \mathcal{N}\left(m + \frac{\sigma^2}{\sigma^2+1}(x_i - m), \frac{\sigma^2}{\sigma^2+1}\right) \end{aligned}$$

Calculando directamente,

$$\begin{aligned} (\hat{\mu}_i - \mu_i)^2 &= \left((\overline{x} - \mu_i) + \left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x})\right)^2 \\ &= (\overline{x} - \mu_i)^2 + 2(\overline{x} - \mu_i)\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x}) + \left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)^2 (x_i - \overline{x})^2. \end{aligned}$$

Por otro lado,

$$\begin{aligned} \mathbb{E}\left\{(\overline{x} - \mu_i)^2 \mid x\right\} &= \overline{x}^2 - 2\overline{x}\mathbb{E}\{\mu_i|x\} + \mathbb{E}\{\mu_i^2|x\} \\ &= \overline{x}^2 - 2\overline{x}\left(m + \frac{\sigma^2}{\sigma^2+1}(x_i - m)\right) + \left(\left(m + \frac{\sigma^2}{\sigma^2+1}(x_i - m)\right)^2 + \frac{\sigma^2}{\sigma^2+1}\right). \end{aligned}$$

y

$$\begin{aligned} \mathbb{E}\left\{(\overline{x} - \mu_i)\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x}) \mid x\right\} &= \left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x})\left(\overline{x} - \mathbb{E}\{\mu_i|x\}\right) \\ &= \left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x})\left(\overline{x} - \left(m + \frac{\sigma^2}{\sigma^2+1}(x_i - m)\right)\right). \end{aligned}$$

De donde,

$$\mathbb{E}\left\{(\hat{\mu}_i - \mu_i)^2 \mid x\right\} =$$
$$\overline{x}^2 - 2\overline{x}\left(m + \frac{\sigma^2}{\sigma^2+1}(x_i - m)\right) + \left(\left(m + \frac{\sigma^2}{\sigma^2+1}(x_i - m)\right)^2 + \frac{\sigma^2}{\sigma^2+1}\right) +$$
$$2\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x})\left(\overline{x} - \left(m + \frac{\sigma^2}{\sigma^2+1}(x_i - m)\right)\right) +$$
$$\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)^2 (x_i - \overline{x})^2 =$$
$$\overline{x}^2 - 2\overline{x}m - 2\overline{x}\frac{\sigma^2}{\sigma^2+1}(x_i - m) + m^2 + 2m\frac{\sigma^2}{\sigma^2+1}(x_i - m) + \left(\frac{\sigma^2}{\sigma^2+1}\right)^2 (x_i - m)^2 + \frac{\sigma^2}{\sigma^2+1} +$$
$$2\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x})(\overline{x} - m) - 2\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)(x_i - \overline{x})\left(\frac{\sigma^2}{\sigma^2+1}(x_i - m)\right) +$$
$$\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)^2 (x_i - \overline{x})^2.$$

Entonces

$$\sum_{i=1}^{n} \mathbb{E}\left\{(\hat{\mu}_i - \mu_i)^2 \mid x\right\} =$$
$$n\overline{x}^2 - 2n\overline{x}m - 2\overline{x}\frac{\sigma^2}{\sigma^2+1}n(\overline{x} - m) + nm^2 + 2m\frac{\sigma^2}{\sigma^2+1}n(\overline{x} - m) + \left(\frac{\sigma^2}{\sigma^2+1}\right)^2 \|x - m\mathbf{1}\|^2 + n\frac{\sigma^2}{\sigma^2+1} +$$
$$0 - 2\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)\frac{\sigma^2}{\sigma^2+1}\|x - \overline{x}\mathbf{1}\|^2 +$$
$$\left(1 - \frac{n-3}{\|x - \overline{x}\mathbf{1}\|^2}\right)^2 \|x - \overline{x}\mathbf{1}\|^2 =$$
$$n\overline{x}^2 - 2n\overline{x}m - 2\frac{\sigma^2}{\sigma^2+1}n(\overline{x}^2 - m\overline{x}) + nm^2 +$$
$$2m\frac{\sigma^2}{\sigma^2+1}n(\overline{x} - m) + \left(\frac{\sigma^2}{\sigma^2+1}\right)^2 \sum_{i=1}^{n}(x_i - m)^2 + n\frac{\sigma^2}{\sigma^2+1} +$$
$$(-2)\frac{\sigma^2}{\sigma^2+1}\|x - \overline{x}\mathbf{1}\|^2 + 2(n-3)\frac{\sigma^2}{\sigma^2+1} +$$
$$\|x - \overline{x}\mathbf{1}\|^2 - 2(n-3) + \frac{(n-3)^2}{\|x - \overline{x}\mathbf{1}\|^2}$$

donde ocupamos que

$$\sum_{i=1}^{n}(x_i - m) = n(\overline{x} - m),$$
$$\sum_{i=1}^{n}(x_i - \overline{x})(x_i - m) = \left(\sum_{i=1}^{n} x_i^2\right) - n\overline{x}^2 = \sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x} + \overline{x}\right) = \|x - \overline{x}\mathbf{1}\|^2,$$
$$\sum_{i=1}^{n} C(x_i - \overline{x}) = C(n\overline{x} - n\overline{x}) = 0 \qquad \forall C \in \mathbb{R} \text{ que no depende de } i.$$

Por lo tanto,

$$\mathbb{E}\left\{\|\hat{\mu}-\mu\|^2\right\} =$$

$$\mathbb{E}\left\{\sum_{i=1}^{n}\mathbb{E}\left\{(\hat{\mu}_i-\mu_i)^2 \mid x\right\}\right\} =$$

$$n\left(\frac{\sigma^2+1}{n}+m^2\right)-2nm^2-2\frac{\sigma^2}{\sigma^2+1}n\left(\left(\frac{\sigma^2+1}{n}+m^2\right)-m^2\right)+nm^2+$$

$$2m\frac{\sigma^2}{\sigma^2+1}n(m-m)+\left(\frac{\sigma^2}{\sigma^2+1}\right)^2 n(\sigma^2+1)+n\frac{\sigma^2}{\sigma^2+1}+$$

$$(-2)\frac{\sigma^2}{\sigma^2+1}(\sigma^2+1)(n-1)+2(n-3)\frac{\sigma^2}{\sigma^2+1}+$$

$$(\sigma^2+1)(n-1)-2(n-3)+\frac{n-3}{\sigma^2+1} =$$

$$\sigma^2+1+nm^2-2nm^2-2\sigma^2+nm^2+$$

$$0+\frac{\sigma^4}{\sigma^2+1}n+n\frac{\sigma^2}{\sigma^2+1}+$$

$$(-2)\sigma^2(n-1)+2(n-3)\frac{\sigma^2}{\sigma^2+1}+$$

$$(\sigma^2+1)(n-1)-2(n-3)+\frac{n-3}{\sigma^2+1} =$$

$$1-\sigma^2+$$

$$n\frac{\sigma^2}{\sigma^2+1}(\sigma^2+1)+$$

$$(n-1)\left(\sigma^2+1-2\sigma^2\right)+$$

$$(n-3)\left(2\left(\frac{\sigma^2}{\sigma^2+1}-1\right)+\frac{1}{\sigma^2+1}\right) =$$

$$1-\sigma^2+$$

$$n\sigma^2+$$

$$(n-1)(1-\sigma^2)+$$

$$(n-3)\left(-\frac{1}{\sigma^2+1}\right) =$$

$$1-\sigma^2+n\sigma^2+(n-n\sigma^2-1+\sigma^2)-\frac{n-3}{\sigma^2+1} =$$

$$n-\frac{n-3}{\sigma^2+1} =$$

$$\frac{n\sigma^2+3}{\sigma^2+1}$$

donde ocupamos que

$$\begin{aligned}
\mathbb{E}\{\overline{x}^2\} &= \mathrm{Var}\{\overline{x}\} + (\mathbb{E}\{\overline{x}\})^2 = \frac{\sigma^2+1}{n} + m^2, \\
\mathbb{E}\left\{\sum_{i=1}^n (x_i - m)^2\right\} = \sum_{i=1}^n \mathbb{E}\left\{(x_i-m)^2\right\} &= \sum_{i=1}^n \mathrm{Var}\{x_i\} = n(\sigma^2+1), \\
\mathbb{E}\{\|x - \overline{x}\mathbf{1}\|^2\} &= (\sigma^2+1)(n-1), \\
\mathbb{E}\left\{\frac{n-3}{\|x-\overline{x}\mathbf{1}\|^2}\right\} &= \frac{1}{\sigma^2+1}.
\end{aligned}$$

La tercera igualdad se cumple porque $\frac{1}{\sigma^2+1}\|x-\overline{x}\mathbf{1}\|^2 \sim \chi^2_{n-1}$ (pues $x \sim \mathcal{N}(m, \sigma^2+1)$) y la cuarta igualdad se cumple porque si $Z \sim \chi^2_k$, entonces $\mathbb{E}\{1/Z\} = \frac{1}{k-2}$. $\square$

Lema. (Stein) Sea $x \sim \mathcal{N}(\mu, \sigma^2)$ y f diferenciable tal que $\mathbb{E}(f(x)(x-\mu))$ y $\mathbb{E}(f'(x))$ existen. Entonces

$$\mathbb{E}(f(x)(x-\mu)) = \sigma^2 \mathbb{E}(f'(x)).$$

2. Sean $\mu_1, \ldots, \mu_n \in \mathbb{R}$, $\sigma^2 > 0$, y $\nu \in \mathbb{R}^n$ (con $n \geq 3$) fijos y arbitrarios. Si

$$x_i \sim \mathcal{N}(\mu_i, \sigma^2), \qquad i = 1, \ldots, n,$$

y

$$\hat{\mu}^{(JS)} := \left(1 - \frac{(n-2)\sigma^2}{\|x-\nu\|^2}\right)(x-\nu) + \nu, \qquad n \geq 3.$$

donde $x = (x_1, \ldots, x_n)$, entonces

$$\mathbb{E}\left\{\|\hat{\mu}^{(JS)} - \mu\|^2\right\} < \mathbb{E}\left\{\|\hat{\mu}^{(ML)} - \mu\|^2\right\}.$$

Es decir, demuestra el teorema de James-Stein para cualquier varianza σ^2 y valor para centrar ν.

Demostración. Calculando directamente,

$$\mathbb{E}\left\{\|\hat{\mu}^{(ML)}-\mu\|^2-\|\hat{\mu}^{(JS)}-\mu\|^2\right\}=$$

$$\mathbb{E}\left\{\|x-\mu\|^2-\left\|\left(1-\frac{(n-2)\sigma^2}{\|x-\nu\|^2}\right)(x-\nu)+\nu-\mu\right\|^2\right\}=$$

$$\mathbb{E}\left\{\|x-\mu\|^2-\left\|(x-\mu)-\left(\frac{(n-2)\sigma^2}{\|x-\nu\|^2}\right)(x-\nu)\right\|^2\right\}=$$

$$\mathbb{E}\left\{\sum_{i=1}^n(x_i-\mu_i)^2-\sum_{i=1}^n\left((x_i-\mu_i)-\left(\frac{(n-2)\sigma^2}{\|x-\nu\|^2}\right)(x_i-\nu_i)\right)^2\right\}=$$

$$\mathbb{E}\left\{\sum_{i=1}^n(x_i-\mu_i)^2-\sum_{i=1}^n\left((x_i-\mu_i)^2-2(x_i-\mu_i)\left(\frac{(n-2)\sigma^2}{\|x-\nu\|^2}\right)(x_i-\nu_i)+\left(\frac{(n-2)^2\sigma^4}{\|x-\nu\|^4}\right)(x_i-\nu_i)^2\right)\right\}=$$

$$\mathbb{E}\left\{\sum_{i=1}^n\left(2(x_i-\mu_i)\left(\frac{(n-2)\sigma^2}{\|x-\nu\|^2}\right)(x_i-\nu_i)-\left(\frac{(n-2)^2\sigma^4}{\|x-\nu\|^4}\right)(x_i-\nu_i)^2\right)\right\}=$$

$$\mathbb{E}\left\{2(n-2)\sigma^2\sum_{i=1}^n\frac{(x_i-\nu_i)}{\|x-\nu\|^2}(x_i-\mu_i)-\left(\frac{(n-2)^2\sigma^4}{\|x-\nu\|^4}\right)\sum_{i=1}^n(x_i-\nu_i)^2\right\}=$$

$$2(n-2)\sigma^2\sum_{i=1}^n\mathbb{E}\left\{\frac{(x_i-\nu_i)}{\|x-\nu\|^2}(x_i-\mu_i)\right\}-\mathbb{E}\left\{\frac{(n-2)^2\sigma^4}{\|x-\nu\|^2}\right\}.$$

Por otro lado, sea

$$f_i(x_i)=\frac{x_i-\nu_i}{\|x-\nu\|^2},\qquad i=1,\dots,n.$$

Entonces

$$f_i'(x_i)=\frac{1}{\|x-\nu\|^2}-\frac{2(x_i-\nu_i)^2}{\|x-\nu\|^4}$$

y

$$\begin{aligned}
\mathbb{E}\left\{\|\hat{\mu}^{(ML)}-\mu\|^2-\|\hat{\mu}^{(JS)}-\mu\|^2\right\}&=2(n-2)\sigma^2\sum_{i=1}^n\mathbb{E}\left\{f_i(x_i)(x_i-\mu_i)\right\}-\mathbb{E}\left\{\frac{(n-2)^2\sigma^4}{\|x-\nu\|^2}\right\}\\
&=2(n-2)\sigma^2\sum_{i=1}^n\sigma^2\mathbb{E}\left\{f_i'(x_i)\right\}-\mathbb{E}\left\{\frac{(n-2)^2\sigma^4}{\|x-\nu\|^2}\right\}\\
&=2(n-2)\sigma^4\sum_{i=1}^n\mathbb{E}\left\{\frac{1}{\|x-\nu\|^2}-\frac{2(x_i-\nu_i)^2}{\|x-\nu\|^4}\right\}-\mathbb{E}\left\{\frac{(n-2)^2\sigma^4}{\|x-\nu\|^2}\right\}\\
&=2(n-2)\sigma^4\mathbb{E}\left\{\frac{n}{\|x-\nu\|^2}-\frac{2\|x-\nu\|^2}{\|x-\nu\|^4}\right\}-\mathbb{E}\left\{\frac{(n-2)^2\sigma^4}{\|x-\nu\|^2}\right\}\\
&=2(n-2)\sigma^4\mathbb{E}\left\{\frac{n-2}{\|x-\nu\|^2}\right\}-\mathbb{E}\left\{\frac{(n-2)^2\sigma^4}{\|x-\nu\|^2}\right\}\\
&=\mathbb{E}\left\{\frac{(n-2)^2\sigma^4}{\|x-\nu\|^2}\right\}\\
&>0.
\end{aligned}$$

Por lo tanto,

$$\mathbb{E}\left\{\|\hat{\mu}^{(JS)}-\mu\|^2\right\} < \mathbb{E}\left\{\|\hat{\mu}^{(ML)}-\mu\|^2\right\}.$$

$\square$

3. Demuestre que el teorema de James-Stein se satisface cuando se tienen m observaciones $x^j = (x_1^j, \ldots, x_n^j)$ con las cuales el estimador de James-Stein esta dado por

$$\hat{\mu}^{(JS)} = \left(1 - \frac{(n-2)\frac{\sigma^2}{m}}{\|\overline{x}\|^2}\right)\overline{x}$$

donde $\overline{x} = \frac{1}{m}(\sum_{j=1}^m x_1^j, \ldots, \sum_{j=1}^m x_n^j)$.

Demostración. Sea $i = 1, \ldots, n$ fijo. Como $x_i^j \sim \mathcal{N}(\mu_i, \sigma^2)$ para todo $j = 1, \ldots, m$ y las variables aleatorias $x_i^1, \ldots, x_i^m$ son independientes, entonces

$$\overline{x}_i = \frac{1}{m}\sum_{j=1}^m x_i^j \sim \mathcal{N}(\mu_i, \sigma^2/m). \qquad \text{(cf. Casella, Berger; Theorem 5.3.1)}$$

Usando esto y el ejercicio 2 obtenemos lo deseado. $\square$

4. En el problema de regresión de cresta supón que $\overline{X} = 0$, es decir, que los covariables se han centrado. Encuentre el mínimo de

$$J_1(w, w_0) = (y - Xw - w_0\mathbf{1})'(y - Xw - w_0\mathbf{1}) + \lambda w'w,$$
$$J_2(w, w_0) = (y - Xw - w_0\mathbf{1})'(y - Xw - w_0\mathbf{1}) + \lambda(w'w + w_0^2).$$

Demostración. Como $\overline{X} = 0$, $\sum_{i=1}^n X_{ij} = 0$ para todo $j = 1, \ldots, p$. Además, calculando directamente,

$$\begin{aligned}
\frac{\partial}{\partial w_k}\sum_{i=1}^n (y_i - (Xw)_i - w_0)^2 &= \sum_{i=1}^n \frac{\partial}{\partial w_k}\left(y_i - \sum_{j=1}^p X_{ij}w_j - w_0\right)^2 \\
&= \sum_{i=1}^n 2\left(y_i - \sum_{j=1}^p X_{ij}w_j - w_0\right)(-X_{ik}) \\
&= 2\sum_{i=1}^n\left[-X_{ik}y_i + \sum_{j=1}^p X_{ik}X_{ij}w_j + w_0X_{ik}\right] \\
&= -2\sum_{i=1}^n X'_{ki}y_i + 2\sum_{j=1}^p\left(\sum_{i=1}^n X'_{ki}X_{ij}\right)w_j + 2w_0\sum_{i=1}^n X_{ik} \\
&= -2(X'y)_k + 2((X'X)w)_k + 0.
\end{aligned}$$

De donde,

$$\frac{\partial}{\partial w}(y - Xw - w_0\mathbf{1})'(y - Xw - w_0\mathbf{1}) = -2X'y + 2X'Xw.$$

Entonces

$$\frac{\partial}{\partial w}J_i(w, w_0) = -2X'y + 2X'Xw + 2\lambda w, \qquad i = 1, 2.$$

En particular,

$$w = (X'X + \lambda I)^{-1} X'y \qquad \text{si} \quad \frac{\partial}{\partial w} J_i(w, w_0) = 0, \qquad i = 1, 2.$$

Por otro lado,

$$\begin{aligned}
\frac{\partial}{\partial w_0} \sum_{i=1}^{n} (y_i - (Xw)_i - w_0)^2 &= \sum_{i=1}^{n} \frac{\partial}{\partial w_0} \left(y_i - \sum_{j=1}^{p} X_{ij} w_j - w_0 \right)^2 \\
&= \sum_{i=1}^{n} 2 \left(y_i - \sum_{j=1}^{p} X_{ij} w_j - w_0 \right) (-1) \\
&= -2 \sum_{i=1}^{n} y_i + 2 \sum_{j=1}^{p} \left(\sum_{i=1}^{n} X_{ij} \right) w_j + 2n w_0 \\
&= -2 \sum_{i=1}^{n} y_i + 2n w_0.
\end{aligned}$$

Entonces

$$\begin{aligned}
\frac{\partial}{\partial w_0} J_1(w, w_0) &= -2 \sum_{i=1}^{n} y_i + 2n w_0, \\
\frac{\partial}{\partial w_0} J_2(w, w_0) &= -2 \sum_{i=1}^{n} y_i + 2n w_0 + 2\lambda w_0.
\end{aligned}$$

En particular,

$$\begin{aligned}
w_0 &= \frac{\sum_{i=1}^{n} y_i}{n} \qquad && \text{si} \quad \frac{\partial}{\partial w} J_1(w, w_0) = 0, \\
w_0 &= \frac{\sum_{i=1}^{n} y_i}{n + \lambda} \qquad && \text{si} \quad \frac{\partial}{\partial w} J_2(w, w_0) = 0.
\end{aligned}$$

□

5. Cuando $p >> n$ en la regresión ridge, utilizamos la descomposición $X = UDV'$ donde V es $(p \times p)$ ortogonal, D es $(p \times p)$ diagonal, y U es $(n \times p)$-dimensional tal que $U'U = I_p$. Muestre que con $R = UD$,

$$\hat{\beta}^{(\text{RIDGE})} = V(D^2 + \lambda I_n)^{-1} R'y.$$

Demostración. Calculando directamente,

$$\begin{aligned}
\hat{\beta}^{(\text{RIDGE})} &= (X'X + \lambda I_n)^{-1} X'y \\
&= \left((UDV')'(UDV') + \lambda I_n \right)^{-1} (UDV')'y \\
&= \left((VDU')(UDV') + \lambda I_n \right)^{-1} VR'y \\
&= \left(VD^2V' + \lambda I_n \right)^{-1} VR'y \\
&= \left(V(D^2V' + \lambda V')^{-1} \right)^{-1} VR'y \\
&= (D^2V' + \lambda V')^{-1} V^{-1} VR'y \\
&= \left((D^2 + \lambda I_n)V' \right)^{-1} R'y \\
&= (V')^{-1} (D^2 + \lambda I_n)^{-1} R'y \\
&= V(D^2 + \lambda I_n)^{-1} R'y.
\end{aligned}$$

□