



CIMAT

Centro de Investigación en Matemáticas
Unidad Monterrey

Análisis Multimodal
Tarea 2

Gustavo Hernández Angeles

15 de noviembre de 2025

Índice

1 Ejercicio 1:	3
1.1 Inciso a)	3
1.2 Inciso b)	3
2 Ejercicio 2:	5

Ejercicio 1:

Utilizando el conjunto de datos `College` disponible en la librería `ISLR`, predice el número de solicitudes recibidas (`Apps`) utilizando las otras variables del conjunto de datos.

- a) Divide el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba.
- b) Ajusta un modelo lineal utilizando mínimos cuadrados en el conjunto de entrenamiento y reporta el error de prueba obtenido.
- c) Ajusta un modelo de regresión ridge en el conjunto de entrenamiento, con λ elegido por validación cruzada. Reporta el error de prueba obtenido.
- d) Ajusta un modelo Lasso en el conjunto de entrenamiento, con λ elegido por validación cruzada. Reporta el error de prueba obtenido, junto con el número de estimaciones de coeficientes distintos de cero.
- e) Ajusta un modelo PCR en el conjunto de entrenamiento, con M elegido por validación cruzada. Reporta el error de prueba obtenido, junto con el valor de M seleccionado por validación cruzada.
- f) Ajusta un modelo PLS en el conjunto de entrenamiento, con M elegido por validación cruzada. Reporta el error de prueba obtenido, junto con el valor de M seleccionado por validación cruzada.
- g) Comenta los resultados obtenidos. ¿Con qué precisión podemos predecir la cantidad de solicitudes universitarias recibidas? ¿Hay mucha diferencia entre los errores de prueba resultantes de estos cinco enfoques?
- h) Propón un modelo (o un conjunto de modelos) que parezca funcionar bien en este conjunto de datos y justifica tu respuesta. Asegúrate de evaluar el rendimiento del modelo utilizando el error del conjunto de validación, la validación cruzada o alguna otra alternativa razonable, en lugar de utilizar el error de entrenamiento. ¿El modelo que elegiste incluye todas las características del conjunto de datos? ¿Por qué o por qué no?

1.1 Inciso a)

Para este problema se dividió el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba, utilizando el 50% de los datos para entrenamiento y el 50% restante para prueba.

1.2 Inciso b)

Se ajustó un modelo lineal utilizando Mínimos Cuadrados Ordinarios (MCO) en el conjunto de entrenamiento, empleando todos los predictores. Las variables que resultaron ser estadísticamente significativas (con $p < 0.1$) en dicho modelo fueron:

- `F.Undergrad` (***)
- `Room.Board` (***)
- `Expend` (**)

- `Grad.Rate` (*)
- `perc.alumni` (.)

mientras que las demás variables no mostraron significancia estadística en el modelo ajustado. Al evaluar el rendimiento de este modelo en el conjunto de prueba, se obtuvo un Error Cuadrático Medio (MSE) de **2,551,734**. Esto corresponde a un Error Cuadrático Medio Raíz (RMSE) de **1,597.4**.

Ejercicio 2:

Es bien sabido que la regresión ridge tiende a dar valores de coeficientes similares a las variables correlacionadas, mientras que lasso puede dar valores de coeficientes totalmente diferentes a las variables correlacionadas. Se explorará esta propiedad en un entorno sencillo.

Supongamos que $n = 2$, $p = 2$, $x_{11} = x_{12}$, $x_{21} = x_{22}$. Además, supongamos que $y_1 + y_2 = 0$ y $x_{12} + x_{22} = 0$, de modo que la estimación del intercepto en mínimos cuadrados, regresión de Ridge o en el modelo de lasso es cero: $\hat{\gamma}_0 = 0$.

- a) Plantea el problema de la optimización con la regresión ridge bajo estas suposiciones.
- b) Argumenta que bajo estas suposiciones, las estimaciones de los coeficientes de ridge satisfacen $\hat{\beta}_1 = \hat{\beta}_2$.
- c) Plantea el problema de la optimización con la regresión lasso bajo estas suposiciones.
- d) Argumenta que en este contexto, los coeficientes de lasso $\hat{\beta}_1$ y $\hat{\beta}_2$ no son únicos; es decir, hay muchas soluciones posibles al problema de optimización en (c). Describe estas soluciones.