

Clase auxiliar # 5

Método LASSO y experimentos

Canción de hoy: Bohemian Rhapsody - Luis Jara.

¡Recordar!

- **Método LASSO:** este método busca minimizar la suma de los residuos al cuadrado (como ya se vio antes), pero penalizada por un factor que permite controlar la cantidad de variables.

$$\min_{\hat{\beta}} \sum_{n=1}^N (Y_n - \hat{\beta}_0 - \sum_{k=1}^K \hat{\beta}_k X_{nk})^2 + \lambda \sum_{k=1}^K |\hat{\beta}_k|$$

- **Efecto de tratamiento y sesgo de selección:**

$$\begin{aligned} & \mathbb{E}(Y_i = 1|D_i = 1) - \mathbb{E}(Y_i = 0|D_i = 0) \\ &= \underbrace{\mathbb{E}(Y_i = 1|D_i = 1) - \mathbb{E}(Y_i = 0|D_i = 1)}_{\text{efecto de tratamiento en tratados}} + \underbrace{\mathbb{E}(Y_i = 0|D_i = 1) - \mathbb{E}(Y_i = 0|D_i = 0)}_{\text{sesgo de selección}} \end{aligned}$$

1. Método LASSO

Tras fracasar en la Liga Kalos, nuestro héroe Ash Ketchum de Pueblo Paleta, ha decidido dedicarse a su verdadera vocación: la política. Afortunadamente y contra todo pronóstico, Ash ha logrado la elección como presidente de su país. Los analistas políticos buscan una explicación y lo contratan a usted para seleccionar aquellas variables que **no** deberían considerar en su análisis.

Para llevar a cabo esta tarea, se le ha entregado la base de datos almacenada en el archivo `auxiliar5.xlsx` (almacenado en Material Docente de U-Cursos). Dicha tiene información de distintas ciudades y pueblos, como el promedio de escolaridad, desempleo, ingreso, nivel de los pokemones de la zona, un índice de ideología política, la popularidad de Pikachu y una serie de “variables basura”.

1. Elija un valor de λ y ejecute el método LASSO.
2. ¿Qué variables descartaría en primera instancia? ¿Por qué?
3. Discuta: ¿cuál es la importancia del parámetro λ en el problema de optimización?

Comente:

4. “El método LASSO sirve para seleccionar variables o forma funcional correcta dentro de un modelo, incluso si hay un número mayor de variables que observaciones”.
5. “No es posible aproximar un modelo cuya forma funcional no sea polinómica mediante el método LASSO”.

2. Experimentos

1. En un experimento para evaluar el programa Chile Solidario, se observa la variable $Y_n = D_n Y_n(1) + (1 - D_n) Y_n(0)$, que es un “índice de bienestar del hogar”.

Por su parte D_n corresponde a una variable dummy que indica si el hogar n recibió la visita de un(a) asistente social. $Y_n(1)$ es el índice del hogar n si recibe el tratamiento e $Y_n(0)$ es el índice si es que no lo recibe.

Explique por qué un diseño experimental que asigna aleatoriamente tratamientos a los hogares permitiría estimar el efecto del tratamiento en el grupo tratado.

2. Un amigo suyo le muestra la siguiente tabla, contenida en una revista de salud, donde se exhibe un indicador de salud promedio de gente dentro y fuera de un hospital¹.

	Personas dentro del hospital	Personas fuera del hospital
Indicador de salud promedio	3	6

Sorprendido por los resultados, su amigo exclama “¡yo siempre supe que ir al hospital es perjudicial para la salud!”. ¿Qué error está cometiendo su amigo?

3. Indignado con la obstinación de su amigo, usted decide realizar un experimento aleatorizado² y obtiene la siguiente tabla.

	Personas sanas	Personas enfermas
Personas que fueron al hospital	7	5
Personas que no fueron al hospital	6	3

Señale los valores de: (1) la diferencia entre enfermos tratados y sanos no tratados, (2) el sesgo de selección y (3) el efecto del tratamiento para personas tratadas.

¹Este indicador de salud es 1 si la salud es “muy mala” y 7 si es “muy buena”.

²Considerar que en la vida real esto es poco factible en un tema tan delicado como la salud de las personas