

## 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:**

$$\mathbb{E}(Y_i = 1 | D_i = 1) - \mathbb{E}(Y_i = 0 | D_i = 0)$$

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

## 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  $\lambda$  y ejecute el método LASSO.

Ver archivo Auxiliar 05 Excel LASSO.xlsx.

2. ¿Qué variables descartaría en primera instancia? ¿Por qué?

Descartaría aquellas variables cuyo coeficiente estimado se hace cero<sup>1</sup> tras la optimización del método LASSO. Dichas variables coinciden con aquellas que denominamos “variables basura” en el enunciado.

3. Discuta: ¿cuál es la importancia del parámetro  $\lambda$  en el problema de optimización?

La importancia del parámetro  $\lambda$  radica en que este representa la magnitud de una penalización marginal en la función objetivo sobre el término  $\sum_{k=1}^K |\hat{\beta}_k|$ . Cuando este término crece en una unidad, la función objetivo –que queremos minimizar– crece  $\lambda$ .

A mayores valores de  $\lambda$ , la función objetivo se verá más castigada por el exceso de variables en el modelo y la optimización será más rigurosa al momento de seleccionar variables. Por lo tanto,  $\lambda$  reflejará lo selectivos que estamos siendo al controlar la cantidad de variables del modelo.

**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”.

**Verdadero.** El método LASSO sirve para seleccionar un número pequeño de variables dentro de un grupo de  $K$  regresores donde  $K$  puede ser mayor al tamaño muestral  $N$ .

5. “No es posible aproximar un modelo cuya forma funcional no sea polinómica mediante el método LASSO”.

**Falso.** Cualquier forma funcional puede ser aproximada por polinomios y productos de polinomios de las variables del modelo, por lo cual es posible aproximar la forma funcional también por LASSO, al incluir polinomios y producto de polinomios en la lista de  $K$  regresores entre los cuales el método escogería aquellos que minimicen su función objetivo.

## 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.

---

<sup>1</sup>Al hacer la optimización en *Excel*, los coeficientes no son exactamente cero, pero se acercan bastante.

En primer lugar, plantearemos el siguiente modelo para estimar el impacto  $\beta_1$  de la visita del asistente social sobre el índice de bienestar del hogar:

$$Y_n = \beta_0 + \beta_1 D_n + U_n$$

Notemos que:

- $\mathbb{E}(Y_n(1)|D_n = 1) = \beta_0 + \beta_1$  (valor esperado del índice para hogares que fueron tratados dado que pertenecen al grupo de tratamiento).
- $\mathbb{E}(Y_n(0)|D_n = 1) = \beta_0$  (valor esperado del índice para hogares que **no** fueron tratados dado que pertenecen al grupo de control).

Esto implica que el efecto que se estimará será la diferencia entre tratados pertenecientes al grupo seleccionado para el tratamiento y no tratados pertenecientes al grupo seleccionado para el control.

$$\beta_1 = \mathbb{E}(Y_n(1)|D_n = 1) - \mathbb{E}(Y_n(0)|D_n = 0)$$

Sin embargo esta diferencia se puede explicar tanto al tratamiento (que puede ser efectivo o no) como a la selección (que puede ser aleatoria o no). Para entender esto, haremos un “ni quita ni pone” con el término  $\mathbb{E}(Y_i = 0|D_i = 1)$  (valor esperado del índice para hogares pertenecientes al grupo seleccionado para el tratamiento suponiendo que no fueron tratados). Claramente este valor no es observable.

$$\beta_1 = \underbrace{\mathbb{E}(Y_i = 1|D_i = 1) - \mathbb{E}(Y_i = 0|D_i = 1)}_{\text{efecto de tratamiento en tratados}} + \underbrace{\mathbb{E}(Y_i = 0|D_i = 1) - \mathbb{E}(Y_i = 0|D_i = 0)}_{\text{sesgo de selección}}$$

Al asignar aleatoriamente el grupo de control y de tratamiento, los resultados esperados en ambos grupos deben ser iguales.

$$\Rightarrow \mathbb{E}(Y_i = 0|D_i = 1) = \mathbb{E}(Y_i = 0|D_i = 0)$$

$$\Rightarrow \beta_1 = \underbrace{\mathbb{E}(Y_i = 1|D_i = 1) - \mathbb{E}(Y_i = 0|D_i = 1)}_{\text{efecto de tratamiento en tratados}}$$

Al seleccionar aleatoriamente los grupos de control y de tratamiento, estimar  $\beta_1$  será estimar el efecto de tratamiento en tratados, sin sesgo de selección de por medio.

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 hospital<sup>2</sup>.

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

<sup>2</sup>Este indicador de salud es 1 si la salud es “muy mala” y 7 si es “muy buena”.

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?

La diferencia entre gente tratada dentro de un hospital y gente no tratada fuera de un hospital se puede descomponer en un efecto de tratamiento y un sesgo de selección (diferencia sistemática entre gente que está dentro y fuera del hospital, que ignora el efecto del tratamiento).

Su amigo está omitiendo la existencia de un sesgo de selección. Claramente el hecho de estar o no en el hospital no es un criterio aleatorio de selección.

3. Indignado con la obstinación de su amigo, usted decide realizar un experimento aleatorizado<sup>3</sup> 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: (a) la diferencia entre enfermos tratados y sanos no tratados, (b) el sesgo de selección y (c) el efecto del tratamiento para personas tratadas.

a)  $\mathbb{E}(Y_n(1)|D_n = 1) - \mathbb{E}(Y_n(0)|D_n = 0) = 5 - 6 = -1$

b)  $\mathbb{E}(Y_n(0)|D_n = 1) - \mathbb{E}(Y_n(0)|D_n = 0) = 3 - 6 = -3$

c)  $\mathbb{E}(Y_n(1)|D_n = 1) - \mathbb{E}(Y_n(0)|D_n = 1) = 5 - 3 = 2$

Notemos que la diferencia entre enfermos tratados y sanos no tratados es la suma entre el sesgo de selección y el efecto de tratamiento:  $(-1) = (-3) + 2$ .

---

<sup>3</sup>Considerar que en la vida real esto es poco factible en un tema tan delicado como la salud de las personas.