Métodos Computacionales 2023
Trabajo Práctico 1:
Regresión Lineal Múltiple

Valentino Pivotto
Luciano Silva Davidov

2 Ejercicios

Primera parte. El objetivo de esta sección es deducir una fórmula para la solución óptima β^* siguiendo los pasos a continuación:

a) Mostrar que el espacio columna de la matriz X es un subespacio vectorial de $\mathbb{R}^n$:

$$Col(X) = \{b \text{ en } \mathbb{R}^n \text{ tales que } b = X\beta \text{ con } \beta \text{ variando en } \mathbb{R}^p\}$$

Subespacio - Definición
Un subespacio de $\mathbb{R}^n$ es cualquier conjunto H en $\mathbb{R}^n$ que cumpla con las siguientes 3 propiedades:
- El vector cero pertenece a H
- Para cada $\mathbf{u}$ y $\mathbf{v}$ en H, la suma $\mathbf{u}+\mathbf{v}$ está en H
- Para cada $\mathbf{u}$ en H y cada escalar c, el vector $c\mathbf{u}$ está en H

Luego, aplicando esta definición con $H = Col(X)$:

$$Col(X) = Gen\{\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_p\}$$

- El vector cero pertenece a $Col(X)$ ya que

$$\mathbf{0} = 0 * \mathbf{x}_1 + 0 * \mathbf{x}_2 + \cdots + 0 * \mathbf{x}_p$$

- Para cada $\mathbf{u}$ y $\mathbf{v}$ en $Col(X)$, la suma $\mathbf{u}+\mathbf{v}$ está en $Col(X)$ puesto que existen $i,\ j$ tal que:

$$\mathbf{u} = s_1 \cdot \mathbf{x}_1 + s_2 \cdot \mathbf{x}_2 + \cdots + s_p \cdot \mathbf{x}_p$$
$$\mathbf{v} = t_1 \cdot \mathbf{x}_1 + t_2 \cdot \mathbf{x}_2 + \cdots + t_p \cdot \mathbf{x}_p$$
$$\mathbf{u}+\mathbf{v} = (s_1 + t_1) \cdot \mathbf{x}_1 + (s_2 + t_2) \cdot \mathbf{x}_2 + \cdots + (s_p + t_p) \cdot \mathbf{x}_p$$

Como es posible escribir al vector $\mathbf{u}$ como una combinación lineal de columnas de X (pesadas con factores s), y también es posible escirbir a $\mathbf{v}$ de forma análoga (es decir, pesado con factores t), la suma de ellos necesariamente pertenece también a $Col(X)$.

- Para cada $\mathbf{u}$ en $Col(X)$ y cada escalar c, el vector $c\mathbf{u}$ está en $Col(X)$ pues:

$$c\mathbf{u} = c \cdot (s_1 \cdot \mathbf{x}_1 + s_2 \cdot \mathbf{x}_2 + \cdots + s_p \cdot \mathbf{x}_p)$$
$$c\mathbf{u} = (c \cdot s_1) \cdot \mathbf{x}_1 + (c \cdot s_2) \cdot \mathbf{x}_2 + \cdots + (c \cdot s_p) \cdot \mathbf{x}_p$$

b) Supongamos que cuando hablamos de vectores en $\mathbb{R}^n$ nos referimos a vectores columna de $\mathbb{R}^{n\times 1}$. Mostrar en ese caso que el producto escalar entre dos vectores $\mathbf{u}$, $\mathbf{v}$ en $\mathbb{R}^n$ puede calcularse como:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^T\mathbf{u}$$

donde la operación en el lado derecho de la igualdad es el producto de matrices usual.

Aplicando la definición de producto matricial:

$$\sum_{i=1}^{n} \mathbf{u}_{i1} \cdot \mathbf{v}_{i1} = \sum_{i=1}^{n} \mathbf{v}_{1i}^{T} \cdot \mathbf{u}_{i1}$$
$$\sum_{i=1}^{n} \mathbf{u}_{i1} \cdot \mathbf{v}_{i1} - \sum_{i=1}^{n} \mathbf{v}_{1i}^{T} \cdot \mathbf{u}_{i1} = 0$$
$$\sum_{i=1}^{n} \mathbf{u}_{i1} \cdot \mathbf{v}_{i1} - \mathbf{v}_{1i}^{T} \cdot \mathbf{u}_{i1} = 0$$
$$\sum_{i=1}^{n} \mathbf{u}_{i1}(\mathbf{v}_{i1} - \mathbf{v}_{1i}^{T}) = 0$$

Tal igualdad implica dos casos disjuntos:

- $\sum_{i=1}^{n} \mathbf{u}_{i1} = 0 \ \forall\ 1 \leq i \leq n$ Este caso no va a suceder siempre
- $\sum_{i=1}^{n} \mathbf{v}_{i1} - \mathbf{v}_{1i}^{T} = 0 \ \forall\ 1 \leq i \leq n$ Este caso sucederá siempre pues:

$$\sum_{i=1}^{n} \mathbf{v}_{i1} - \mathbf{v}_{1i}^{T} = 0$$
$$\sum_{i=1}^{n} \mathbf{v}_{i1} = \mathbf{v}_{1i}^{T}$$

Lo cual vale $\forall$ vector $\mathbf{v}$ en $\mathbb{R}^n$

c) Aplicando el teorema tomando como subespacio S el subespacio del ítem (a), el punto y de $\mathbb{R}^n$ como el vector de la variable dependiente, y el vector b como $b = X\beta^*$, convertir esta ecuación de optimalidad

$$\parallel y - X\beta^* \parallel = \min_{\beta\ en\ \mathbb{R}^p} \parallel y - X\beta \parallel$$

en la condición de ortogonalidad que corresponde a la equivalencia 2 del teorema.

Teorema
Sea y un vector cualquiera de $\mathbb{R}^n$ y S un subespacio de $\mathbb{R}^n$. El vector de S que minimiza la distancia del subespacio S al vector y es aquel b de S tal que $y - b$ es ortogonal a todo vector s de S. Es decir, las siguientes dos condiciones son equivalentes:
1. $\parallel y - b \parallel = \min_{s \in S} \parallel y - s \parallel$
2. $(y - b) \cdot s = 0$ para todo s en S

Aplicando este teorema con $S = Col(X)$, el cual demostramos es un subespacio de $\mathbb{R}^n$, y con $b = X\beta^*$

$$\parallel y - X\beta^* \parallel = \min_{s\ en\ Col(X)} \parallel y - s \parallel$$

$$\parallel y - X\beta^* \parallel = \min_{s\ en\ \{b \text{ en } \mathbb{R}^n \text{ tales que } b = X\beta \text{ con } \beta \text{ variando en } \mathbb{R}^p\}} \parallel y - s \parallel$$

Como X es siempre la misma matriz, entonces lo único que va a variar es β, y sabemos que β varía en $\mathbb{R}^p$. También sabemos que $b = X\beta$, entonces hacemos los reemplazos:

$$\parallel y - X\beta^* \parallel = \min_{\beta\ en\ \mathbb{R}^p} \parallel y - X\beta \parallel$$

Luego, por teorema, esta condición es equivalente a la de ortogonalidad, es decir:

$$(y - X\beta^*) \cdot X\beta = 0 \text{ para todo } \beta \text{ en } \mathbb{R}^p$$

d) A la ecuación obtenida en el ítem (c), aplicarle la identidad del producto escalar vista en el item (b), para llegar a la ecuación:

$$X^T(y - X\beta^*) \cdot \beta = \mathbf{0}$$

Partiendo de la ecuación obtenida en el ítem (c):

$$(y - X\beta^*) \cdot X\beta = 0 \text{ para todo } \beta \text{ en } \mathbb{R}^p$$

Recordando la propiedad del producto escalar demostrada en (b): $\mathbf{u} \cdot \mathbf{v} = \mathbf{v}^T\mathbf{u}$ donde $\mathbf{u}$, $\mathbf{v}$ son vectores en $\mathbb{R}^n$, es posible emplearla tomando los siguientes parámetros:

$\mathbf{u} = (y - X\beta^*)$ pues $\in \mathbb{R}^n$ ya que $y \in \mathbb{R}^n$ y $X \in \mathbb{R}^{n\times p}$, $\beta^* \in \mathbb{R}^p$ por lo que $X\beta^* \in \mathbb{R}^n$

$\mathbf{v} = X\beta$ pues $\in \mathbb{R}^n$ dado que $X \in \mathbb{R}^{n\times p}$ y $\beta \in \mathbb{R}^p$
Así:

$$(X\beta)^T(y - X\beta^*) = 0$$

Distribución de la transposición

$$\beta^T X^T(y - X\beta^*) = 0$$

Ahora es posible volver a emplear la propiedad de (b) pero en sentido inverso ($\mathbf{v}^T\mathbf{u} = \mathbf{u}\cdot\mathbf{v}$), con estos parámetros y en $\mathbb{R}^p$ en lugar de $\mathbb{R}^n$:
$\mathbf{v}^T = \beta^T$ pues $\in \mathbb{R}^p$ ya que $\beta \in \mathbb{R}^p$
$\mathbf{u} = X^T(y - X\beta^*)$ pues $\in \mathbb{R}^p$ dado que $X^T \in \mathbb{R}^{p\times n}$ y $(y - X\beta^*) \in \mathbb{R}^n$
De esta manera:

$$X^T(y - X\beta^*) \cdot \beta = 0$$

e) Se sabe que el único vector que es ortogonal a todo vector $\mathbf{v}$ de $\mathbb{R}^n$ es el vector nulo. Es decir, si $\mathbf{u}$ es un vector fijo tal que $\mathbf{u} \cdot \mathbf{v} = \mathbf{0}$ para todo $\mathbf{v}$ en $\mathbb{R}^n$, entonces $\mathbf{u} = \mathbf{0}$. Usando esto y la ecuación obtenida en el ítem (d), llegar a la fórmula:

$$X^T X\beta^* = X^T y$$

Partiendo de la ecuación obtenida en el ítem (d):

$$X^T(y - X\beta^*) \cdot \beta = 0$$

Distribución de X^T

$$(X^T y - X^T X\beta^*) \cdot \beta = 0$$

Esta igualdad se satisface si y solo si se cumple al menos uno de estos casos:
Caso 1: $(X^T y - X^T X\beta^*) = 0$
Caso 2: $\beta = 0$
De la definición de $Col(X)$ provista en el inciso (a) se obtiene que β varía en $\mathbb{R}^p$, es por esto

que β no puede ser $\mathbf{u}$ en la propiedad del enunciado. Si β no es $\mathbf{u}$, entonces $\beta \neq \mathbf{0}$, por lo que solo el caso 1 puede ser válido.
Luego, como el caso 1 es el válido:

$$X^T y - X^T X \beta^* = 0$$
$$X^T y = X^T X \beta^*$$

f) Finalmente, suponiendo que las columnas de X son linealmente independientes, se tiene que la matriz $X^T X$ es invertible. Despejar β^* de la ecuación del ítem (e) para llegar a la fórmula de la solución óptima al problema de regresión.

$$X^T X \beta^* = X^T y$$

Multiplicación a izquierda por $(X^T X)^{-1}$ en ambos lados de la igualdad

$$(X^T X)^{-1} X^T X \beta^* = (X^T X)^{-1} X^T y$$

$$I \beta^* = (X^T X)^{-1} X^T y$$

$$\beta^* = (X^T X)^{-1} X^T y$$