

# Clase 16. Matrices de Proyección y Método de Mínimos Cuadrados.

Curso ‘Linear Algebra’ del MIT.

## Resumen

Continuaremos estudiando la matriz de proyección  $P$  vista en la clase anterior, usada para encontrar la mejor solución a un sistema de ecuaciones lineales irresoluble. Posteriormente, nos centraremos en buscar la línea de mejor ajuste de datos a partir del método de mínimos cuadrados.

## 16.1 Matrices de Proyección (recapitulación).

Como recordaremos de la clase anterior, cuando tenemos un sistema  $A\vec{x} = \vec{b}$  con  $\vec{b} \notin C(A)$  (i.e, sin solución), buscamos un vector  $\vec{p} \in C(A)$  que se proyecte sobre  $\vec{b}$ , el cual usamos para resolver  $A\hat{x} = \vec{p}$  donde  $\hat{x}$  es la mejor solución posible para  $A\vec{x} = \vec{b}$ .

También vimos que el vector  $\hat{x}$  lo encontrábamos por medio de la siguiente fórmula:

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

En ese sentido,  $A\hat{x} = \vec{p}$  podemos expresarlo como:

$$A \left[ (A^T A)^{-1} A^T \vec{b} \right] = \vec{p}$$

Si reordenamos esta ecuación como:

$$\left[ A(A^T A)^{-1} A^T \right] \vec{b} = \vec{p}$$

se da que  $P = A(A^T A)^{-1} A^T$  es la matriz de proyección con la cual obtenemos  $\vec{p}$ . Por lo tanto,

reemplazando en la ecuación de arriba:

$$P\vec{b} = \vec{p}$$

De esta ecuación podemos sacar las siguientes conclusiones: Primero, si  $\vec{b} \in C(A)$ , entonces  $P\vec{b} = \vec{b}$  puesto que se cumple la igualdad  $A\vec{x} = \vec{b}$  debido a aquella condición. Esto podemos demostrarlo multiplicando la fórmula de  $P$  por  $\vec{b}$ .

$$P\vec{b} = (A(A^T A)^{-1} A^T) \vec{b} = (A(A^T A)^{-1} A^T) (A\vec{x}) = A((A^T A)^{-1} A^T A) \vec{x} = A I \vec{x} = A\vec{x} = \vec{b}$$

Y, en segundo lugar, si  $\vec{b} \perp C(A)$  entonces  $P\vec{b} = \vec{0}$ , porque significa que  $\vec{b} \in N(A^T)$  y sabemos que este subespacio es un complemento ortogonal del  $C(A)$ . Como consecuencia,  $A^T \vec{b} = \vec{0}$ . Demostremoslo multiplicando a  $P$  por  $\vec{b}$  otra vez:

$$P\vec{b} = (A(A^T A)^{-1} A^T) \vec{b} = (A^T A)^{-1} (A^T \vec{b}) = (A^T A)^{-1} \vec{0} = \vec{0}$$

### 16.1.1 Proyección de $\vec{e}$ sobre $\vec{b}$ .

Recordemos que al proyectar  $\vec{p}$  sobre  $\vec{b}$ , se genera un vector  $\vec{e} = \vec{b} - \vec{p}$  que es ortogonal a  $\vec{p}$  y, por consiguiente, al  $C(A)$  debido a que  $\vec{p} \in C(A)$ . Por lo tanto,  $\vec{e} \in N(A^T)$  implicando que  $A^T \vec{e} = \vec{0}$ .

En ese sentido, a  $\vec{e}$  podemos entenderlo como un **vector de proyección de  $\vec{b}$ , pero que es perpendicular a  $\vec{p}$** , el cual también tiene su matriz de proyección.

Sean  $\vec{e} = \vec{b} - \vec{p}$  y  $\vec{p} = P\vec{b}$ . Entonces:

$$\vec{e} = \vec{b} - P\vec{b} = \vec{b}(I - P)$$

donde  $(I - P)$  es la **Matriz de Proyección** con la cual obtenemos a  $\vec{e}$  y cumple las mismas características que  $P$ :

$$(I - P^T) = (I - P) \quad \text{y} \quad (I - P^2) = (I - P)$$

Por otra parte, algo muy relevante que no está demás señalar es que, al sumar a  $\vec{e}$  con  $\vec{p}$ , obtenemos al vector  $\vec{b}$ :

$$\vec{b} = \vec{e} + \vec{p}$$

Es relevante no solo como una consecuencia algebraica a partir de  $\vec{e}$ , sino también porque

ahora sabemos obtener tanto a este vector como a  $\vec{p}$  a partir de sus matrices de proyección. Por lo tanto, al obtenerlas podemos buscar ambos vectores para, posteriormente, tener al  $\vec{b}$ .

## 16.2 Método de Mínimos Cuadrados.

La clase pasada vimos una aplicación de la proyección sobre un subespacio vectorial, conocido como el **método de mínimos cuadrados** que consiste en buscar la recta de mejor ajuste sobre un conjunto de datos.

Cuando tenemos dos variables  $x$  e  $y$ , con  $y$  dependiente de  $x$ , cuyos datos siguen una relación aparentemente lineal, podría interesarnos generar un modelo que nos permita **predecir** valores de salida (de  $y$ ) futuros debido a este patrón detectado. Es decir, de la forma:

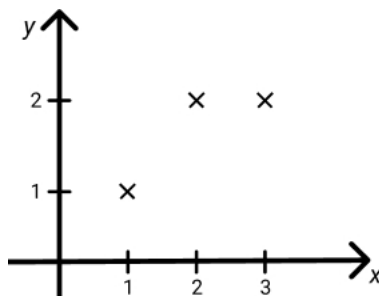
$$y = mx + b$$

con  $m$  siendo la pendiente y  $b$  la intersección en el eje vertical o el valor de  $y$  cuando  $x = 0$ .

El problema es que los datos casi nunca siguen una tendencia perfectamente lineal. Por ejemplo, sea el siguiente conjunto:

$x$	1	2	3
$y$	1	2	2

y su gráfico:



En el método de mínimos cuadrados se propone encontrar una recta que pase lo más cerca de los puntos del gráfico. Por ello, el interés se centra en **buscar los mejores valores para la pendiente  $m$  y la intersección  $b$**  tal que el ajuste sea el más apropiado a los datos. Escribamos la ecuación como suele hacerse en este método, donde  $\beta_0 = b$  y  $\beta_1 = m$ .

$$y = \beta_0 + \beta_1 x$$

Ahora, cuando reemplazamos los valores de  $x$  e  $y$  de la tabla en la ecuación de arriba, nos encontramos con la siguiente dificultad:

$$\begin{cases} \beta_0 + \beta_1 1 = 1 \\ \beta_0 + \beta_1 2 = 2 \\ \beta_0 + \beta_1 3 = 2 \end{cases}$$

Como vemos, terminamos con un **sistema sobredeterminado**, que es cuando tiene más ecuaciones que incógnitas y suele ser inconsistente o no tener soluciones. Esto podemos verlo pasándolo a matrices.

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \quad \vec{x} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \vec{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

La matriz de coeficientes  $A$  es de rango columna completo,  $\text{rango}(A) = 2$ . Como vimos en la Clase 8, esto significa que tiene una o ninguna solución. Para ver si este  $\vec{b}$  es el correcto, podemos reescribirlo como  $\vec{b} = \langle b_1, b_2, b_3 \rangle$ . Luego, escribir a  $A$  como una matriz aumentada y llevarla a una escalada reducida.

$$[A|\vec{b}] = \left[ \begin{array}{cc|c} 1 & 1 & b_1 \\ 1 & 2 & b_2 \\ 1 & 3 & b_3 \end{array} \right] \Rightarrow \left[ \begin{array}{cc|c} 1 & 1 & b_1 \\ 0 & 1 & b_2 - b_1 \\ 0 & 2 & b_3 - b_1 \end{array} \right] \Rightarrow \left[ \begin{array}{cc|c} 1 & 0 & 2b_1 - b_2 \\ 0 & 1 & b_2 - b_1 \\ 0 & 0 & b_3 - 3b_1 - 2b_2 \end{array} \right]$$

El vector  $\vec{b}$  correcto debe ser aquel donde en sus componentes se cumpla que  $b_3 - 3b_1 - 2b_2 = 0$ . Sin embargo, aquello no se cumple con el nuestro:

$$2 - 3(1) - 2(2) = -5 \neq 0$$

Por lo tanto, ninguna combinación lineal entre  $\vec{x} = \langle \beta_0, \beta_1 \rangle$  y los vectores columna de  $A$  será igual a  $\vec{b} = \langle 1, 1, 2 \rangle$ . Es decir,  $\vec{b} \notin C(A)$ , implicando que  $A\vec{x} = \vec{b}$  no tiene una solución<sup>1</sup>.

El método de mínimos cuadrados en Álgebra Lineal consiste en proyectar un vector  $\vec{p} \in C(A)$  sobre  $\vec{b} \notin C(A)$  para resolver  $A\hat{x} = \vec{p}$ , donde  $\hat{x}$  es el vector de los mejores  $\beta_0$  y  $\beta_1$  para los datos que tenemos.

Ahora que tenemos en cuenta que usaremos una proyección para obtener la recta de mejor ajuste, podemos entender por qué este método se llama “mínimos cuadrados”. Asumiendo que  $\vec{p} \neq \vec{b}$ , la diferencia entre los componentes de ambos vectores que se llama **error** y se

---

<sup>1</sup>También lo vimos en la Clase 8.

denota como  $e$ .

En particular, si  $\vec{p} = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}^T$  y  $\vec{b} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^T$ , entonces:

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} - \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

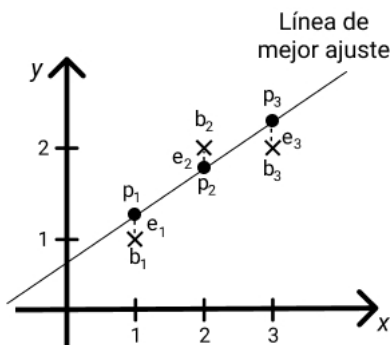
$$\vec{b} - \vec{p} = \vec{e}$$

donde  $\vec{e}$  es el **vector error**.

Este método se llama “de **mínimos cuadrados**” porque su objetivo es **reducir lo más posible la magnitud** (o **longitud**)  $||\vec{e}||$ , la cual también elevamos al cuadrado:

$$||\vec{e}||^2 = ||\vec{b} - \vec{p}||^2$$

Todo esto lo podemos lograr resolviendo  $A\hat{x} = \vec{p}$ , porque  $\vec{p}$  será el más cercano a  $\vec{b}$  o donde se cumplirá que  $||\vec{e}||^2$  será el mínimo. Como resultado, obtendremos una recta que pasará muy cerca de los valores contenidos en  $\vec{b}$  como la que vemos a continuación:



Dicha **minimización** estará reflejada en  $\hat{x} = \langle \hat{\beta}_0, \hat{\beta}_1 \rangle$ , la mejor estimación de  $\vec{x} = \langle \beta_0, \beta_1 \rangle$ . Para ello, multiplicamos a  $A^T$  a  $A\vec{x} = \vec{b}$

$$A^T A \hat{x} = A^T \vec{b}$$

obteniendo las llamadas **Ecuaciones Normales**, que son las que minimizan  $||\vec{e}||^2$ .

Comencemos calculando la matriz  $A^T A$ :

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

Luego sigamos con el vector columna  $A^T \vec{b}$ :

$$A^T \vec{b} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

Por lo tanto:

$$A^T A \hat{x} = A^T \vec{b}$$

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

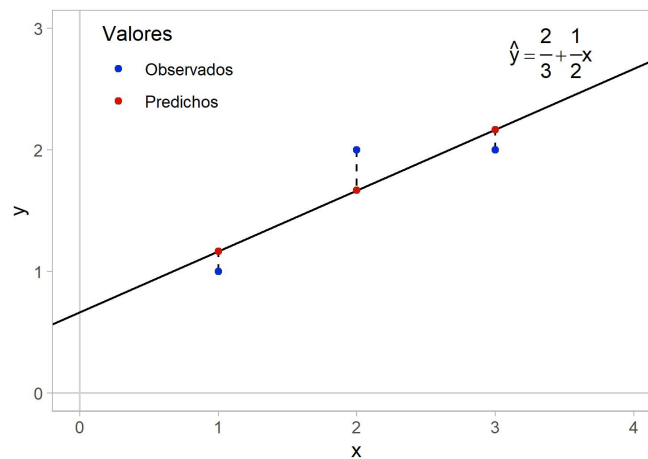
las que corresponden a las siguientes ecuaciones normales:

$$\begin{cases} 3\hat{\beta}_0 + 6\hat{\beta}_1 = 5 \\ 6\hat{\beta}_0 + 14\hat{\beta}_1 = 11 \end{cases}$$

Al resolver este sistema, obtenemos que  $\hat{\beta}_0 = 2/3$  y  $\hat{\beta}_1 = 1/2$ . Por consiguiente, la recta de mejor ajuste está dada por la siguiente ecuación:

$$\hat{y} = \frac{2}{3} + \frac{1}{2}x$$

donde  $\hat{y} = \vec{p}$ , es decir, los valores estimados del modelo lineal. A continuación tenemos su gráfica real:



Los valores de  $\hat{x}$  podríamos haberlos encontrado mucho más rápido, porque la matriz  $A^T A$  de este caso en particular, es de  $\text{rango}(A^T A) = 2$ . En otras palabras, es de rango completo,

lo que significa que tiene una inversa y corresponde a:

$$(A^T A)^{-1} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix}$$

Por lo tanto, al multiplicar  $(A^T A)^{-1}$  a  $A^T A \hat{x} = A^T \vec{b}$  resulta en:

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

Reemplacemos  $(A^T A)^{-1}$  y  $A^T \vec{b}$  con los objetos que manejamos, obteniendo que:

$$\hat{x} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} 5 \\ 11 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

A continuación tenemos una tabla con los valores de  $x$ ,  $y$  (componentes de  $\vec{b}$ ),  $\hat{y}$  (componentes de  $\vec{p}$ ) y de  $e$  (componentes de  $\vec{e}$ ).

$x$	1	2	3
$y$	1	2	2
$\hat{y}$	7/6	5/3	13/6
$e$	-1/6	1/3	-1/6

A partir de esta tabla, veamos si lo que hemos estudiado sobre proyecciones se cumple. En primer lugar, hemos señalado que el vector  $\vec{b}$  es igual a la suma vectorial de  $\vec{p}$  y  $\vec{e}$ . A continuación vemos con los datos que manejamos, que es así:

$$\vec{b} = \vec{p} + \vec{e}$$

$$\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 7/6 \\ 5/3 \\ 13/6 \end{bmatrix} + \begin{bmatrix} -1/6 \\ 1/3 \\ -1/6 \end{bmatrix}$$

También hemos estudiado que  $\vec{e} \perp \vec{p}$ . En otras palabras, debería cumplirse que:

$$\vec{p}^T \cdot \vec{e} = 0$$

$$\begin{bmatrix} 7/6 & 5/3 & 13/6 \end{bmatrix} \cdot \begin{bmatrix} -1/6 \\ 1/3 \\ -1/6 \end{bmatrix} = 0$$

$$0 = 0$$

Y también se cumplirá que el producto punto entre cualquier combinación lineal entre las columnas de  $A$  y  $\vec{e}$  será igual a cero. De hecho, lo acabamos de ver con  $\vec{p}$  que es parte del  $C(A)$ .

## 16.3 Vectores Ortonormales (preliminar).

En la próxima clase profundizaremos en los llamados **vectores ortonormales**.

En la sección anterior vimos que  $A^T A$  era invertible. En la clase pasada vimos que esto es posible solo si **las columnas de  $A$  son linealmente independientes** o, dicho de otro modo, que  $A$  sea de **rango columna completo**. Demostremos esta afirmación.

**Demostración.** Digamos que  $A^T A \vec{x} = \vec{0}$ . La matriz  $A^T A$  debe ser de rango completo para que sea invertible. De ser el caso, debería tener una única solución:  $\vec{x} = \vec{0}$  o que el  $N(A^T A) = \{\vec{0}\}$ .

Comencemos calculando con  $\vec{x}^T$  un producto punto en ambos lados de  $A^T A \vec{x} = \vec{0}$ .

$$\vec{x}^T A^T A \vec{x} = \vec{x}^T \vec{0}$$

$$\vec{x}^T A^T A \vec{x} = 0$$

La ecuación de arriba podemos reordenarla como:

$$(\vec{x}^T A^T)(A \vec{x}) = 0$$

$$(A \vec{x})^T (A \vec{x}) = 0$$

El producto punto  $(A \vec{x})^T (A \vec{x}) = \|A \vec{x}\|^2$ . Por lo tanto, al calcular  $\sqrt{\cdot}$  en ambos lados:

$$\sqrt{\|A \vec{x}\|^2} = \sqrt{0}$$

$$\|A \vec{x}\| = 0$$

Es decir, la magnitud del vector  $A \vec{x}$  es cero y esto solo ocurre en el  $\vec{0}$ . En consecuencia:

$$A \vec{x} = \vec{0}$$

que es equivalente a que el  $N(A) = \{\vec{0}\}$  y esto se cumple solo cuando  $A$  es de rango columna completo o de rango completo.

Supongamos que  $A$  es de rango columna completo. De igual manera  $A^T A$  será de rango completo, porque será cuadrada y todas sus columnas serán linealmente independientes. Por



lo tanto, de aquí surgen dos conclusiones:

1.  $A^T A \vec{x} = \vec{0}$  tiene una sola solución: el vector  $\vec{0}$ .
2.  $A^T A$  es **invertible**.

De hecho, cuando  $A^T A$  tiene una inversa,  $A^T A \vec{x} = \vec{0}$  tiene una solución:

$$\begin{aligned}(A^T A)^{-1} A^T A \vec{x} &= (A^T A)^{-1} \vec{0} \\ \vec{x} &= \vec{0}\end{aligned}$$

Por lo tanto, queda demostrado (Q.E.D) que cuando las columnas de  $A$  son linealmente independientes,  $A^T A$  es invertible. Además, que en ese caso,  $N(A) = N(A^T A) = \{\vec{0}\}$ .

Cuando, por ejemplo, usamos el método de mínimos cuadrados, lo ideal es que  $A^T A$  sea invertible y eso es posible cuando los vectores columnas de  $A$  son **linealmente independientes**. Una forma de asegurarnos que **siempre lo sean**, es que **las columnas sean vectores unitarios<sup>2</sup> perpendiculares entre sí**, los cuales reciben el nombre de **vectores ortonormales**.

Es decir, **si todas las columnas de  $A$  son vectores ortonormales,  $A$  siempre será de rango fila completo**.

---

<sup>2</sup>Vectores de magnitud igual a 1.