UC | Chile

UC | Chile

Algebra Lineal Aplicada para Ciencia de Datos

Clase 9. Problemas cuadráticos: Mínimos cuadrados ordinarios

UC | Chile

Motivación

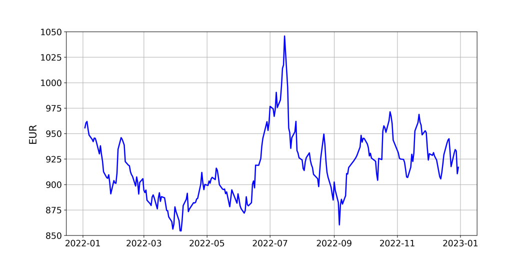
Motivación

Un problema clásico consiste en observar datos

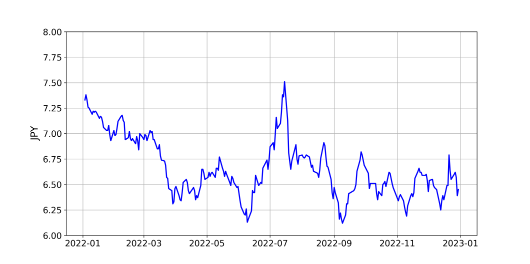
$$(x_1, y_1), \quad \ldots, \quad (x_m, y_m)$$

donde $x_1, \ldots, x_m$ son observaciones del **predictor** y $y_1, \ldots, y_m$ son observaciones de la **respuesta**

Motivación

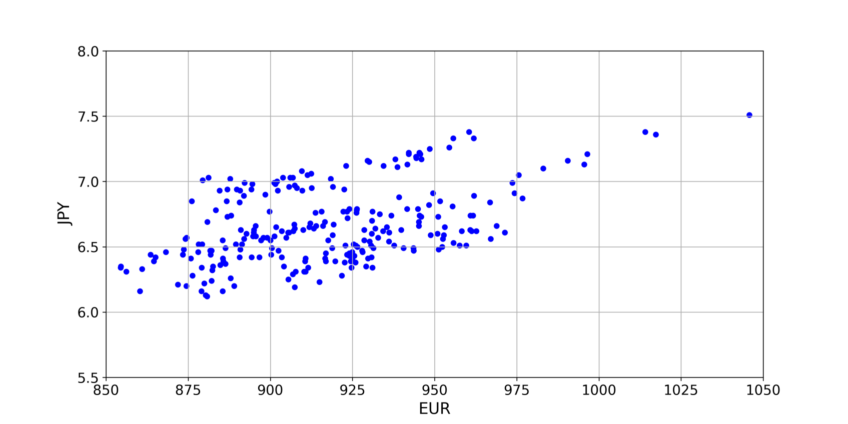


Serie EUR 2022



Serie JPY 2022

Motivación



JPY vs EUR

Motivación

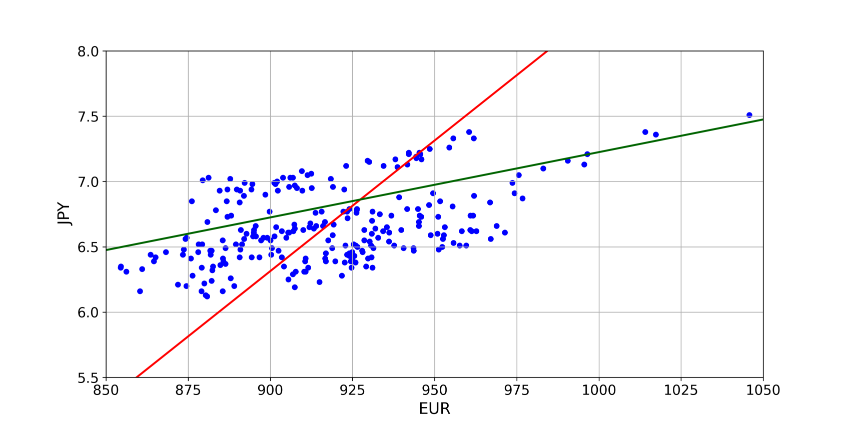
Es de interés **modelar** la dependencia entre el predictor y la respuesta

Un ejemplo de modelo simple es el **modelo afín**

$$y_i \approx \beta_0 + \beta_1 x_i \quad \text{para} \quad i \in \{1, \ldots, m\}$$

Para encontrar los valores β_0, β_1 es necesario definir un **criterio de ajuste**

Motivación



JPY vs EUR

Motivación

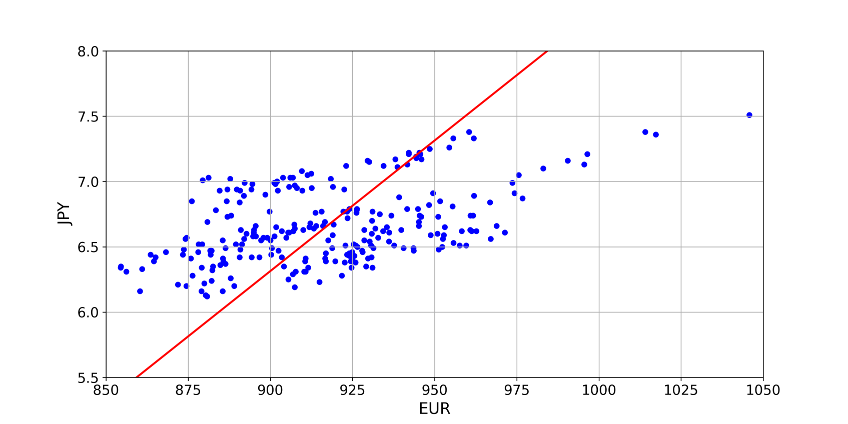
Si definimos el **residuo**

$$r_i = y_i - \beta_0 - \beta_1 x_i \quad \text{para} \quad i \in \{1, \ldots, m\}$$

entonces podemos usar como criterio de ajuste la **suma de los residuos al cuadrado**

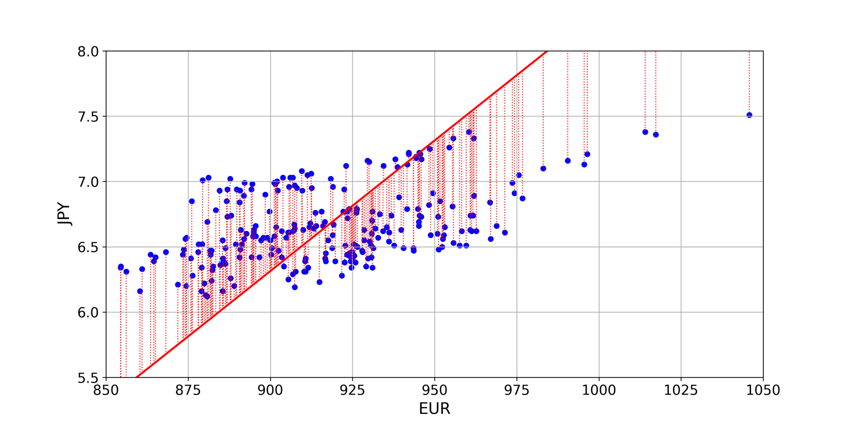
$$\textstyle\sum_{i=1}^{m} r_i^2 = \sum_{i=1}^{m} (y_i - \beta_0 - \beta_1 x_i)^2$$

Motivación



Modelo afín

Motivación



Residuales

Motivación

Si definimos

$$r = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \text{y} \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$$

entonces **matricialmente**

$$\sum_{i=1}^{m} r_i^2 = \|y - X\beta\|_2^2$$

Motivación

Esto nos lleva a resolver el problema de **mínimos cuadrados ordinarios**

$$\underset{\beta \in \mathbb{R}^2}{\text{minimizar}} \quad \|y - X\beta\|_2^2$$

Formulación del problema

Formulación del problema

Consideramos un vector $y \in \mathbb{R}^m$ y una matriz $A \in \mathbb{R}^{m \times n}$ con

$$m > n$$

La matriz es **alta**: tiene **a lo menos** tantas filas como columnas.

Supondremos además que es de **rango completo**: para una matriz alta, esto quiere decir que

$$\mathbf{nul}(A) = \{0\}$$

o bien, que sus columnas son **linealmente independientes**

Formulación del problema

El sistema

$$y = Ax$$

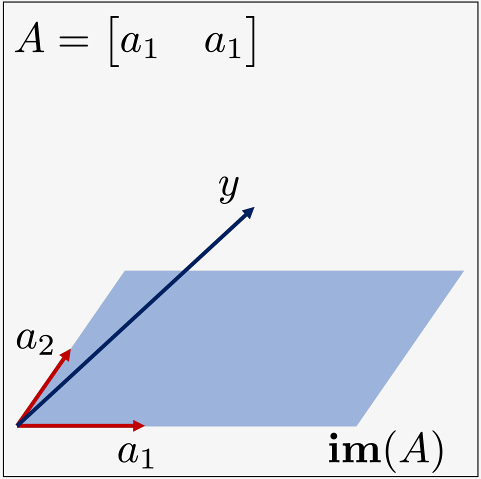
está **sobredeterminado** y puede ser **inconsistente** cuando

$$y \notin \mathbf{im}(A)$$

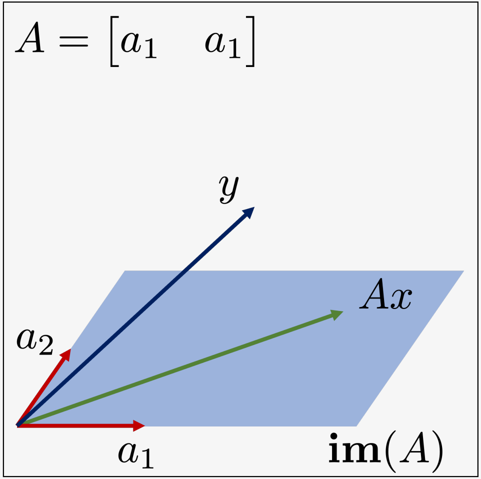
Es por ello que buscamos resolver el problema de **mínimos cuadrados ordinarios**

$$\underset{x \in \mathbb{R}^n}{\text{minimizar}} \quad \|y - Ax\|_2^2$$

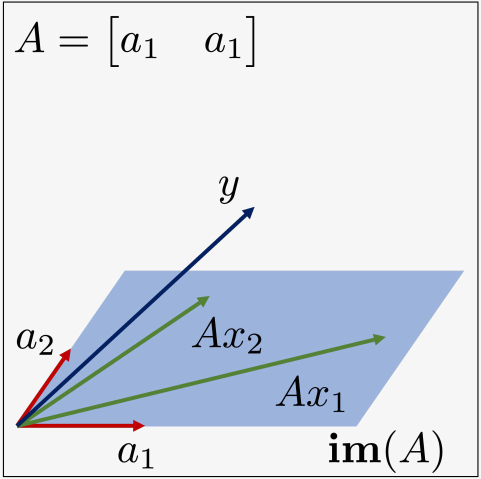
Formulación del problema. Ejemplo



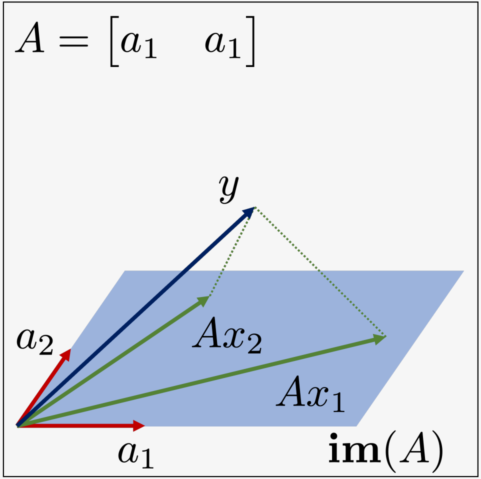
Formulación del problema. Ejemplo



Formulación del problema. Ejemplo



Formulación del problema. Ejemplo



Formulación del problema. Ejemplo

Para

$$y = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \quad \text{y} \quad A = \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 3 & 0 \end{bmatrix}$$

el sistema

$$y = Ax$$

es **inconsistente**

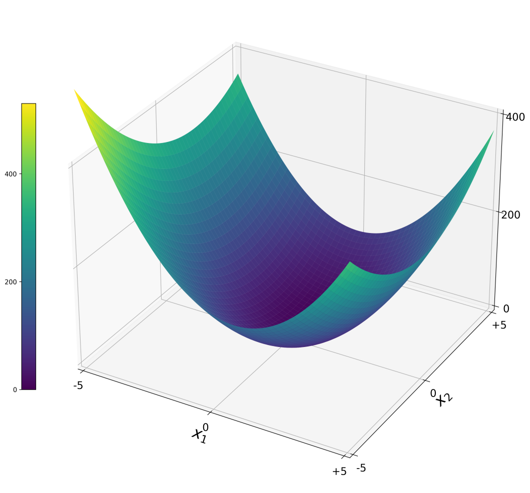
Formulación del problema. Ejemplo

Buscamos un vector x para el cual la suma de residuos al cuadrado

$$\begin{aligned}
\|y - Ax\|_2^2 &= \left\| \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_2^2 \\
&= \left\| \begin{bmatrix} 1 - x_1 + x_2 \\ 3 - x_1 - 2x_2 \\ -3x_1 \end{bmatrix} \right\|_2^2 \\
&= (1 - x_1 + x_2)^2 + (3 - x_1 - 2x_2)^2 + 9x_1^2
\end{aligned}$$

sea mínimo

Formulación del problema. Ejemplo



La geometría de la proyección ortogonal

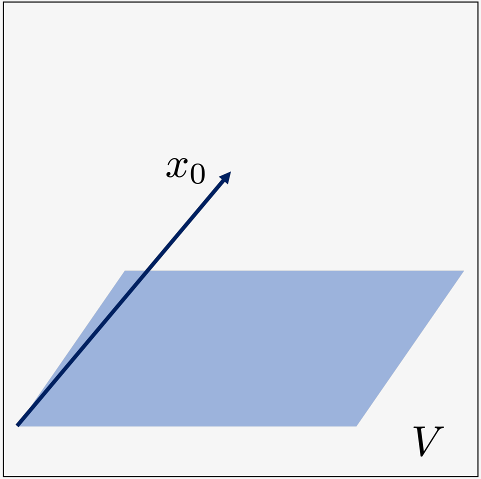
La geometría de la proyección ortogonal

Primero estudiamos **la geometría de la proyección ortogonal**

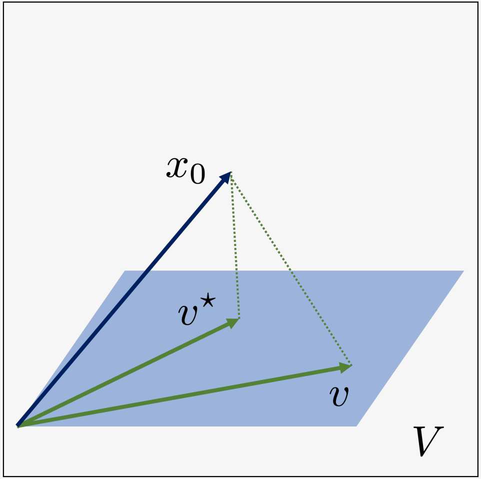
Sea $x_0 \in \mathbb{R}^n$ y $V \subset \mathbb{R}^n$ un **subespacio vectorial** generado por una colección cualquiera de vectores

Buscamos un punto $v^\star \in V$ que esté a la **mínima distancia Euclideana** posible de x_0

La geometría de la proyección ortogonal



La geometría de la proyección ortogonal



La geometría de la proyección ortogonal

El caso más simple ocurre cuando $x_0 \in V$ ya que en tal caso basta con escoger el vector

$$v^\star = x_0$$

que está a distancia 0 de x_0

Por lo tanto, suponemos que $x_0 \notin V$

La geometría de la proyección ortogonal

Si $v^\star$ es el punto a distancia mínima, entonces necesariamente se debe tener que

$$\|x_0 - v^\star\|_2^2 \leq \|x_0 - v\|_2^2$$

para cualquier $v \in V$

La geometría de la proyección ortogonal

Esta condición implica que

$$\begin{aligned}\|x_0 - v^\star\|_2^2 &\leq \|(x_0 - v^\star) - (v - v^\star)\|_2^2 \\ &\leq \|x_0 - v^\star\|_2^2 - 2(x_0 - v^\star) \cdot (v - v^\star) + \|v - v^\star\|_2^2\end{aligned}$$

o bien que

$$(x_0 - v^\star) \cdot (v - v^\star) \leq \frac{1}{2}\|v - v^\star\|_2^2$$

para cualquier $v \in V$

La geometría de la proyección ortogonal

Si existe $z \in V$ con $z \neq 0$ tal que

$$(x_0 - v^\star) \cdot z \neq 0$$

entonces podemos escoger $v = v^\star + \alpha z$ con $\alpha \neq 0$ de modo que

$$(x_0 - v^\star) \cdot (v - v^\star) = \alpha (x_0 - v^\star) \cdot z \leq \frac{\alpha^2}{2} \|z\|_2^2 = \frac{1}{2} \|v - v^\star\|_2^2$$

Esto implica

$$\mathbf{signo}(\alpha)(x_0 - v^\star) \cdot z \leq 0 \quad \Rightarrow \quad (x_0 - v^\star) \cdot z = 0.$$

La geometría de la proyección ortogonal

Ya que $z \in V$ es arbitrario, concluimos que

$$(x_0 - v^{\star}) \cdot z = 0 \quad \text{para cualquier} \quad z \in V.$$

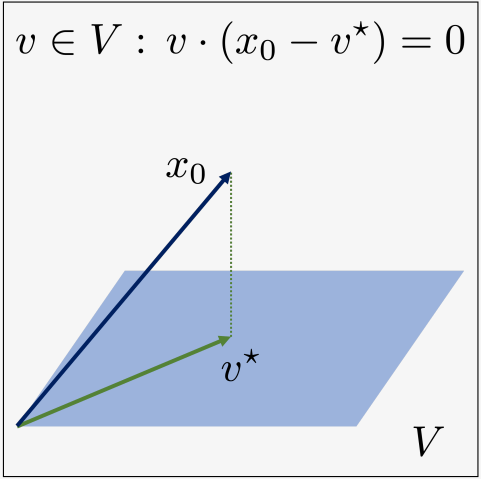
La **proyección ortogonal** de x_0 sobre V es

$$\mathbf{proy}_V(x_0) = v^{\star}.$$

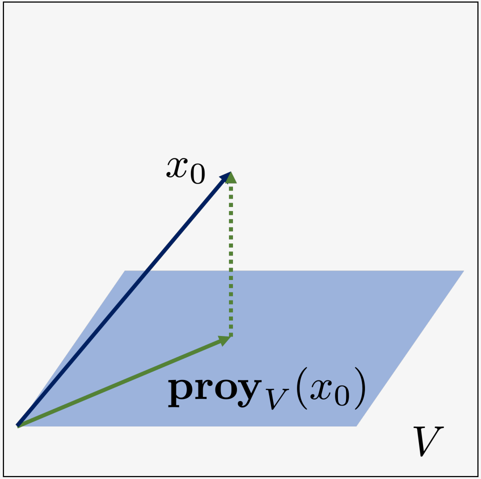
y el **complemento ortogonal** de x_0 sobre V es

$$x_0 - \mathbf{proy}_V(x_0)$$

La geometría de la proyección ortogonal



La geometría de la proyección ortogonal



Resolución con argumentos geométricos

Resolución con argumentos geométricos

$$\boxed{\underset{x \in \mathbb{R}^n}{\text{minimizar}} \quad \|y - Ax\|_2^2}$$

Ya que

$$Ax \in \mathbf{im}(A) \quad \text{para cualquier} \quad x \in \mathbb{R}^n$$

esto es equivalente a

$$\underset{\widehat{y} \in \mathbb{R}^m}{\text{minimizar}} \quad \|y - \widehat{y}\|_2^2 \quad \text{sujeto a} \quad \widehat{y} \in \mathbf{im}(A).$$

Se trata de una proyección ortogonal

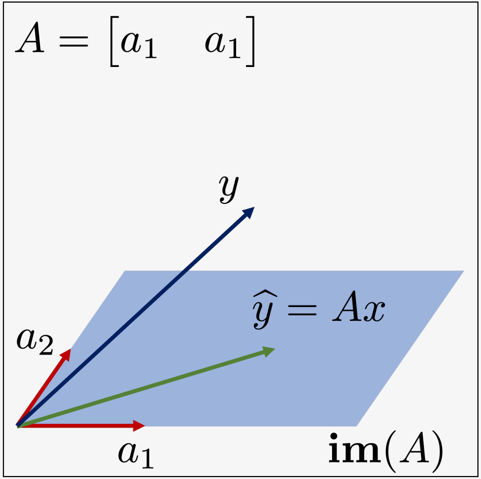
Resolución con argumentos geométricos

$$
\begin{aligned}
\|y - \widehat{y}\|_2^2 &= \|y - \mathbf{proy}_{\mathrm{im}(A)}(y) + \mathbf{proy}_{\mathrm{im}(A)}(y) - \widehat{y}\|_2^2 \\
&= \|y - \mathbf{proy}_{\mathrm{im}(A)}(y)\|_2^2 + 2(y - \mathbf{proy}_{\mathrm{im}(A)}(y)) \cdot (\mathbf{proy}_{\mathrm{im}(A)}(y) - \widehat{y}) \\
&\quad + \|\mathbf{proy}_{\mathrm{im}(A)}(y) - \widehat{y}\|_2^2 \\
&= \|y - \mathbf{proy}_{\mathrm{im}(A)}(y)\|_2^2 + \|\mathbf{proy}_{\mathrm{im}(A)}(y) - \widehat{y}\|_2^2
\end{aligned}
$$

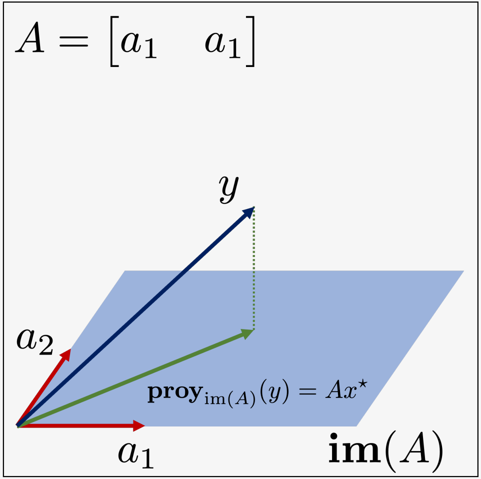
El vector $x^\star$ para el cual se alcanza el mínimo debe resolver la **ecuación lineal**

$$
Ax^\star = \mathbf{proy}_{\mathrm{im}(A)}(y).
$$

La geometría de la proyección ortogonal



La geometría de la proyección ortogonal



Resolución con argumentos geométricos

El vector $x^\star$ para el cual se alcanza el mínimo debe resolver la **ecuación lineal**

$$Ax^\star = \mathbf{proy}_{\mathrm{im}(A)}(y).$$

¿Podemos calcular esto sin calcular la proyección?

Resolución con argumentos geométricos

De la propiedad crucial de la proyección ortogonal

$$\widehat{y} \cdot (y - \mathbf{proy}_{\mathrm{im}(A)}(y)) = 0 \quad \Rightarrow \quad \widehat{y} \cdot y = \widehat{y} \cdot \mathbf{proy}_{\mathrm{im}(A)}(y)$$

para cualquier $\widehat{y} \in \mathbf{im}(A)$ y de

$$A = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} \quad \text{con} \quad a_1, \dots, a_n \in \mathbb{R}^m$$

deducimos que

$$a_i \cdot y = a_i \cdot \mathbf{proy}_{\mathrm{im}(A)}(y) \quad \text{para} \quad i \in \{1, \dots, n\}.$$

Resolución con argumentos geométricos

Ya que

$$A^\top = \begin{bmatrix} a_1^\top \\ \vdots \\ a_n^\top \end{bmatrix}$$

lo anterior implica que

$$A^\top A x^\star = A^\top \mathbf{proy}_{\mathrm{im}(A)}(y) = A^\top y$$

La solución $x^\star$ debe ser solución al sistema de ecuaciones

$$A^\top A x^\star = A^\top y.$$

Las ecuaciones normales

Las ecuaciones normales

La ecuación

$$A^\top A x^\star = A^\top y$$

se conoce como **las ecuaciones normales** o **la ecuación normal**

Es un sistema de ecuaciones en n incógnitas y con n ecuaciones

Ya que A es **alta** y de **rango completo**, la matriz $A^\top A$ es **invertible**

Las ecuaciones normales

Si

$$A^\top A x = 0 \quad \text{entonces} \quad Ax \in \mathbf{nul}(A^\top)$$

Pero

$$\mathbf{nul}(A^\top) = \mathbf{im}(A)^\perp \quad \text{implica} \quad Ax \in \mathbf{im}(A)^\perp$$

Por lo tanto

$$Ax \in \mathbf{im}(A) \cap \mathbf{im}(A)^\perp \quad \Rightarrow \quad Ax = 0$$

Ya que las columnas de A son linealmente independientes, concluimos que $x = 0$

Las ecuaciones normales

De las ecuaciones normales

$$A^\top A x^\star = A^\top y$$

concluimos que

$$x^\star = (A^\top A)^{-1} A^\top y$$

y la **proyección ortogonal de y sobre $\mathbf{im}(A)$** es

$$\widehat{y}^\star = A x^\star = A(A^\top A)^{-1} A^\top y.$$

Esto se interpreta como la **predicción óptima de y**

Las ecuaciones normales. Ejemplo

Para

$$y = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \quad \text{y} \quad A = \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 3 & 0 \end{bmatrix}$$

se tiene

$$A^\top A = \begin{bmatrix} 11 & -3 \\ -3 & 5 \end{bmatrix} \quad \text{y} \quad A^\top y = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$$

y las ecuaciones normales son

$$\begin{bmatrix} 11 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} x_1^\star \\ x_2^\star \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$$

Las ecuaciones normales

Las ecuaciones normales se resuelven en la práctica usando la **factorización QR**

Cuando $A \in \mathbb{R}^{m\times n}$ es **alta** y de **rango completo** entonces

$$A = QR \quad \text{con} \quad Q \in \mathbb{R}^{m\times n} \quad \text{y} \quad Q^\top Q = I \quad \text{y} \quad R \in \mathbb{R}^{n\times n} \quad \text{invertible}$$

Para calcular $x^\star$ resolvemos el sistema

$$A^\top A x^\star = (QR)^\top (QR) x^\star = R^\top Q^\top Q R x^\star = R^\top R x^\star = A^\top y \quad \Rightarrow \quad R^\top R x^\star = A^\top y$$

¿Qué ganamos?

Las ecuaciones normales

Podemos resolver el sistema

$$R^\top R x^\star = A^\top y$$

resolviendo dos **sistemas triangulares**

$$R^\top z^\star = A^\top y$$
$$R x^\star = z^\star$$

usando **sustitución progresiva** y **sustitución regresiva** respectivamente

El caso general

Las ecuaciones normales

¿Qué ocurre cuando A es alta, pero no es de rango completo?

En este caso, las columnas de A son **linealmente dependientes**

Las ecuaciones normales siguen siendo válidas

$$A^\top A x = A^\top y$$

En este caso el sistema es **consistente** pero existe una **cantidad infinita** de soluciones

Si bien existe una infinidad de soluciones a las ecuaciones normales, la predicción óptima está únicamente definida

UC | Chile