Vimos la clase pasada:

<u>Teorema</u>: Dada $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$

son equivalentes:

1) $a^* \in \mathbb{R}^m$ minimiza $\|Ax - b\|_2$

2) a^* es solución de $A^t A x = A^t b$

Además, si las columnas de A son li, la solución a^* de $A^t A x = A^t b$ existe y es única. Notar $A^t A \in \mathbb{R}^{m \times m}$.

Si m es grande la matriz $A^t A$ puede ser enorme y resolver

$A^t A x = A^t b$ muy costoso.

En este caso si hacemos la descomposición QR de A (que asumimos tiene columnas li)

$$\underbrace{\begin{pmatrix} a_1 | \cdots | a_m \end{pmatrix}}_{A \in \mathbb{R}^{n \times m}} = \underbrace{\begin{pmatrix} q_1 | \cdots | q_m \end{pmatrix}}_{Q \in \mathbb{R}^{n \times m}} \underbrace{\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ 0 & r_{22} & & r_{2m} \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & r_{mm} \end{pmatrix}}_{R \in \mathbb{R}^{m \times m}}$$

con $r_{ii} \neq 0$

donde Q es una matriz con columnas que forman un conjunto ortonormal y R es triangular superior inversible.

Entonces hacemos lo siguiente

$$\underset{\mathbb{R}^{n \times m}}{\overset{}{A}} = \underset{\mathbb{R}^{n \times m}}{Q} \cdot \underset{\mathbb{R}^{m \times m}}{R}$$

Q tiene columnas ortonormales con $Q^t \cdot Q = Id$

y R es triangular inversible (los $r_{jj} \neq 0$).

$$\Rightarrow A^t A = R^t Q^t Q R = R^t R$$

$$\Rightarrow A^t A a = A^t b$$

$$\Updownarrow$$

$$R^t R a = R^t Q^t b \quad (R^t \text{ inversible})$$

$$\Downarrow$$

$$R a = Q^t b$$ con R triangular sup

Nota: A^tA crucial pero mala para cómputo. Matrices ortogonales y triangulares son "las buenas".

Pseudo inversa de Penrose (mirar notas clase pasada)

Quedó para probar:

Ejercicio: $A \in \mathbb{k}^{m \times n}$

Si $m > n$ y $rg(A) = n$ (col de A son l.i.)

$$\Rightarrow A^{\dagger} = (A^*A)^{-1}A^* \quad \text{y} \quad A^{\dagger} \cdot A = Id$$

Dem: $A = U \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{pmatrix} V^*$ con $\sigma_n > 0$ ($rg(A) = n$)

$$\Rightarrow A^+ = V \left(\begin{matrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_n^{-1} \end{matrix} \middle| 0 \right) V^*, \quad A^* = V \Sigma^t U^*$$

y $$A^* A = V \Sigma^t \underbrace{U^* U}_{Id} \Sigma V^* \qquad V V^* = V^* V = Id$$
$$= V \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix} V^*$$

$$\Rightarrow (A^* A)^{-1} A^* = V \begin{pmatrix} \sigma_1^{-2} & & \\ & \ddots & \\ & & \sigma_n^{-2} \end{pmatrix} \underbrace{V^* V}_{Id} \left(\begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{matrix} \middle| 0 \right) U^*$$
$$= V \left(\begin{matrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_n \end{matrix} \middle| 0 \right) U^* = A^+$$

luego, si las col de $A \in \mathbb{R}^{m \times n}$ son li, la única sol de $A^t A x = A^t b$ es $x = (A^t A)^{-1} A^t b$
$$= A^+ b.$$

Ej de parcial:

En un bosque se ha identificado una especie de planta exótica. A lo largo de los últimos años se ha registrado cómo progresó el nro. de individuos de dicha especie. Se tienen los datos

año	$\overset{=x_1}{0}$	$\overset{=x_2}{1}$	$\overset{=x_3}{2}$	$\overset{=x_4}{3}$	$\overset{=x_5}{4}$	$\overset{=x_6}{5}$	$\overset{=x_7}{6}$
cant indiv	$\underset{y_1}{216}$	$\underset{y_2}{219}$	$\underset{y_3}{287}$	$\underset{y_4}{303}$	$\underset{y_5}{439}$	$\underset{y_6}{556}$	$\underset{y_7}{681}$

a) Hallar el poli de grado menor o igual a dos q' mejor ajusta a los datos en el sentido de los cuadrados mínimos.

Buscamos $P(x) = a_0 + a_1 x + a_2 x^2$:

$$\sum_{i=1}^{7} \Big(\underbrace{P(x_i) - y_i}_{}\Big)^2 \text{ sea mín.}$$

$$\left(a_0 + a_1 x_i + a_2 x_i^2 - y_i\right)^2$$

Es decir si

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \end{pmatrix} \quad y \quad b = \begin{pmatrix} 216 \\ 219 \\ 287 \\ 303 \\ 439 \\ 556 \\ 681 \end{pmatrix}$$

a_0, a_1, a_2: $\left\| A \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} - b \right\|_2^2$ sea mín.

y vimos que como las columnas de A son l.i (chequear), encontrar $\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$ que haga $\left\| A \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} - b \right\|_2^2$ sea mín equivale a encontrar $\begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$ solución de

$$\underbrace{A^t A}_{3 \times 3} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \underbrace{A^t b}_{3 \times 1}$$

Ej: Programar y resolver en Python.

Si después me piden estimar cuántos individuos habrá en 100 años según este modelo haré $P(100)$.

b) Ajustar una función de la forma $g(x) = d_0 e^{d_1 x}$ aplicando cuadrados mínimos sobre la función transformada convenientemente.

Podríamos plantear la minimización (en función de d_0 y d_1) de

$$\sum_{i=1}^{7} \left(g(x_i) - y_i\right)^2 = \sum_{i=1}^{7} \left(d_0 e^{d_1 x_i} - y_i\right)^2$$

Pero vemos que el parámetro d_1 no es lineal dentro del cuadrado (como lo era para $P(x) = a_0 + a_1 x + a_2 x^2$).

Entonces no podemos usar la técnica anterior en donde reescribimos el problema de minimización equivalente $\left\| A \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} - b \right\|_2$.

Para buscar una aproximación y poder usar los resultados anteriores "linealizamos" nuestro problema tomando ln. (logaritmo neperiano).

No estamos resolviendo el problema

de minimizar $\sum_{i=1}^{7} \left(d_0 e^{d_1 x_i} - y_i\right)^2$

pero resolvemos uno similar:

Buscamos d_0, d_1 que minimice

$$\sum_{i=1}^{7} \left(\ln\left(d_0 e^{d_1 x_i}\right) - \ln y_i \right)^2$$

<u>NotA</u>: Dados $b > a > 1$, por el teorema del valor medio, existe $c \in (a,b)$:

$$\ln b - \ln a = \underbrace{\frac{1}{c}}_{(\ln x)'\big|_{x=c}} (b-a)$$

como $c > 1 \Rightarrow \frac{1}{c} < 1$

$$\Rightarrow |\ln b - \ln a| < |b-a|$$

$$\Rightarrow (\ln b - \ln a)^2 < (b-a)^2$$

De esta manera, si los y_i's de los datos son grandes (como en el ejemplo) es de esperar que si usamos el d_0, d_1 encontrados con la linealización para el problema original, el error sea peor.

Volvamos al planteo de la linealización:

Buscamos d_0, d_1 que minimice

$$\sum_{i=1}^{7} \left(\ln(d_0 e^{d_1 x_i}) - \ln y_i \right)^2 ,$$ es

decir buscamos d_0, d_1 que haga mínimo

$$\sum_{i=1}^{7} \left(\underbrace{\ln(d_0)}_{\substack{\text{lo llamamos} \\ a_0}} + \underbrace{d_1}_{\substack{\text{lo llamamos} \\ a_1}} x_i - \underbrace{\ln y_i}_{\substack{\text{lo llamamos} \\ \tilde{y}_i}} \right)^2$$

este problema equivale a

$$\min_{a_0, a_1} \left\| A \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} - b \right\|_2$$

donde

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{pmatrix} \quad y \quad b = \begin{pmatrix} \ln 216 \\ \ln 219 \\ \ln 287 \\ \ln 303 \\ \ln 439 \\ \ln 556 \\ \ln 681 \end{pmatrix}$$

$\Rightarrow$ planteamos

$$A^t \cdot A \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = A^t b$$

Una vez que resolvemos y encontramos a_0, a_1 $\rightsquigarrow$ volvemos para atrás usando

$$\ln d_0 = a_0 \Rightarrow d_0 = e^{a_0}$$

$$d_1 = a_1$$

c) Para el mismo problema podemos encontrar un polinomio de grado 3, grado 4, etc que mejor aproxime en el sentido de los cuadrados mínimos?

En general, dada una tabla de datos

x	x_1	x_2	$\dots$	x_m
y	y_1	y_2	$\dots$	y_m

se busca un polinomio de grado menor o igual a m que mejor aproxime a mis datos en el sentido de los cuadrados mínimos.

Buscamos $a_0, a_1, \dots, a_m$ ($m+1$ incógnitas)

de manera que si $P(x) = a_0 + a_1 x + \dots + a_m x^m$

$$\sum_{i=1}^{m} \left(P(x_i) - y_i\right)^2 \text{ sea mínimo.}$$

Planteamos

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ \vdots & x_2 & x_2^2 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{pmatrix} \in \mathbb{R}^{n \times m}$$

y $$b = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

El problema es equivalente a buscar un mínimo de

$$\left\| A \begin{pmatrix} a_0 \\ \vdots \\ a_m \end{pmatrix} - b \right\|_2$$

y vimos que si las columnas de A son li $\Rightarrow$ la única solución viene dada por $A^t A \begin{pmatrix} a_0 \\ \vdots \\ a_m \end{pmatrix} = A^t b.$

La pregunta es: cuándo puedo garantizar que los columnas son li?

Para empezar como $A \in \mathbb{R}^{n \times m}$, se tienen $m+1$ columnas que son vectores de $\mathbb{R}^n$.

Si queremos que haya posibilidad de que sean li, debe ser $m+1 \leq n$

Cuando $m+1 = n$

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ \vdots & x_2 & x_2^2 & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix}$$ se conoce como

matriz de Vandermonde y se nota $V[x_1, x_2, \dots, x_n]$.

Veamos que en este caso cuando $V[x_1 \dots x_n]$ es una matriz inversible. Para eso calculemos su determinante por inducción.

$$\det V[x_1,x_2] = \det \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix} = x_2 - x_1$$

Vamos a usar que hacer operaciones de columna $C_i + \alpha C_j \to C_i$ no cambia el determinante de una matriz y que el determinante de multiplicar una columna por un escalar es igual a ese escalar por el determinante.

Es decir, vale que:

- $\det\left(C_1 \,|\, C_2 \,|\, C_i+\alpha C_j \,|\, \dots \,|\, C_n\right) \quad i \neq j$

  $= \det\left(C_1 \,|\, C_2 \,|\, \dots C_i \,|\, \dots \,|\, C_n\right)$

- $\det\left(C_1 \,|\, \,|\alpha C_i| \dots \, C_n\right)$

  $= \det\left(C_1 \,|\, \dots |C_i| \dots |C_n\right)$

Y lo mismo por filas.

$$\det V[x_1\, x_2\, x_3] = \det \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix}$$

$$= \det \begin{pmatrix} 1 & 0 & 0 \\ 1 & x_2 - x_1 & x_2^2 - x_1 x_2 \\ 1 & x_3 - x_1 & x_3^2 - x_1 x_3 \end{pmatrix}$$

$\downarrow$

$C_3 - x_1 C_2 \to C_3$

$C_2 - x_1 C_1 \to C_2$

$$= \det \begin{pmatrix} x_2 - x_1 & x_2^2 - x_1 x_2 \\ x_3 - x_1 & x_3^2 - x_1 x_3 \end{pmatrix}$$

$$= \det \begin{pmatrix} x_2 - x_1 & (x_2 - x_1) x_2 \\ x_3 - x_1 & (x_3 - x_1) x_3 \end{pmatrix}$$

$$= (x_2 - x_1)(x_3 - x_1) \cdot \det \begin{pmatrix} 1 & x_2 \\ 1 & x_3 \end{pmatrix}$$

$$= (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$$

Usando la misma idea para el caso general se prueba que

$$\det V[x_1 \cdots x_n] = \prod_{n \geq j > i \geq 1} (x_j - x_i)$$

Con lo cual vemos que esta matriz es inversible si y solo si $x_i \neq x_j$ $\forall i \neq j$. Es decir los x_i's de la tabla son todos distintos.

Volvamos al caso general con $m \leq n-1$.

Vimos que si $m = n+1$ y $x_i \neq x_j$ $\forall i \neq j$ entonces $V[x_1 \dots x_n]$ es inversible $\Rightarrow$ el rango de la matriz es n, es decir, sus columnas son l.i.

Con lo cual si me quedo con un subconjunto de $m < n-1$ columnas también serán l.i.

En conclusión, dada una tabla de n datos con $x_i \neq x_j$ $\forall i \neq j$, existe un único polinomio de grado menor o igual a m $(\leq n-1)$

que mejor aproxima a la tabla en el sentido de los cuadrados mínimos.

<u>OBS:</u>

- Si los x_i están muy cerca entre sí provoca oscilaciones muy altas
- Si m es muy grande cercano a $n-1$ también produce oscilaciones especialmente cerca de los extremos. (Fenómeno de Runge).

<u>Polinomio interpolador:</u> Con $x_i \neq x_j \ \forall i \neq j$

Para $m = n-1$ se sabe que existe un único polinomio de grado menor o igual a $n-1$ que satisface

$$P(x_i) = y_i \quad i = 1 \dots n.$$

Es decir, este polinomio pasa

por cada puntos (x_i, y_i).

Se lo llama polinomio interpolador de Lagrange.

Podemos encontrarlo como antes planteando $A = V[x_1 \dots x_n]$ y resolviendo

$$A^t A \begin{pmatrix} a_0 \\ \vdots \\ a_{n-1} \end{pmatrix} = A^t \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Si no, podemos usar la base de Lagrange:

Para cada x_j con $j = 1 \dots n$,

definimos $l_j(x)$ el polinomio mónico que satisface:

$$l_j(x_i) = 0 \quad \text{si } i \neq j$$

$$l_j(x_j) = 1,$$

Usamos $x_i \neq x_j$.

es decir $$l_j(x) = \frac{\prod_{i \neq j} (x - x_i)}{\prod_{i \neq j} (x_j - x_i)}$$

$$\Rightarrow \quad P(x) = \sum_{j=1}^{n} y_j \, \ell_j(x)$$

Veamos que P es único. Supongamos existe otro q de grado menor o igual a $n-1$: $q(x_j) = y_j \quad \forall j = 1 \dots n$

$$\Rightarrow \quad (p-q)(x_j) = 0 \quad \forall j = 1 \dots n \quad (*)$$

y $p-q$ tiene grado menor o igual a $n-1$ $\Rightarrow$ tiene a lo sumo $n-1$ raíces. lo cual se contradice con $(*)$.

<u>Error de Interpolación</u>: Sean $f \in C^n[a,b]$ y P el poli interpolador de grado $\leq n-1$ en los puntos $x_1, \dots, x_n$. Para cada $x \in [a,b]$ existe $\xi = \xi(x) \in [a,b]$:

$$E(x) = f(x) - P(x) = \frac{f^{(n)}(\xi)}{n!} \cdot \prod_{i=1}^{n} (x - x_i)$$