

Álgebra Lineal Computacional

Introducción

Segundo Cuatrimestre 2022

Facultad de Ciencias Exactas y Naturales, UBA

¿Por qué Álgebra Lineal?

El álgebra lineal es una de las áreas más hermosas, simples y poderosas de la matemática.

Si escucharon hablar de...

- Big data
- Machine learning
- Aprendizaje supervisado

¿Por qué Álgebra Lineal?

El álgebra lineal es una de las áreas más hermosas, simples y poderosas de la matemática.

Si escucharon hablar de...

- Big data → álgebra lineal
- Machine learning
- Aprendizaje supervisado

¿Por qué Álgebra Lineal?

El álgebra lineal es una de las áreas más hermosas, simples y poderosas de la matemática.

Si escucharon hablar de...

- Big data → álgebra lineal
- Machine learning → álgebra lineal
- Aprendizaje supervisado

¿Por qué Álgebra Lineal?

El álgebra lineal es una de las áreas más hermosas, simples y poderosas de la matemática.

Si escucharon hablar de...

- Big data → álgebra lineal
- Machine learning → álgebra lineal
- Aprendizaje supervisado → álgebra lineal

Ejemplo: el premio Netflix

Conociendo los puntajes de los usuarios de Netflix a distintas películas, poder predecir qué puntajes les van a dar esos usuarios a películas que todavía no vieron.

Y a partir de esas predicciones, hacer recomendaciones.

Fuente: <https://www.netflixprize.com/>

El premio Netflix

Propuesta 1: identificar perfiles de usuarios y asignar a cada usuario a uno de estos perfiles.



El premio Netflix

Propuesta 1: identificar perfiles de usuarios y asignar a cada usuario a uno de estos perfiles.



Herramientas: Aprendizaje no supervisado - Clustering

El premio Netflix

Propuesta 2: Cada persona se compone con un porcentaje de cada uno de esos perfiles.

10%



30%



5%



40%



15%



El premio Netflix

Propuesta 2: Cada persona se compone con un porcentaje de cada uno de esos perfiles.



¡Esto es un sistema lineal de ecuaciones!

El premio Netflix

Cada persona es una combinación lineal de esos perfiles.

0.1 0.3 0.05 0.4 0.15



El premio Netflix

Sabiendo los puntajes que estos usuarios le dieron a a varias películas podemos calcular los coeficientes de la combinación lineal.

$$a_1 \begin{pmatrix} 1 \\ 3 \\ 5 \\ 2 \\ 3 \end{pmatrix} + a_2 \begin{pmatrix} 2 \\ 3 \\ 5 \\ 4 \\ 5 \end{pmatrix} + a_3 \begin{pmatrix} 3 \\ 1 \\ 4 \\ 5 \\ 2 \end{pmatrix} + a_4 \begin{pmatrix} 1 \\ 4 \\ 5 \\ 1 \\ 4 \end{pmatrix} + a_5 \begin{pmatrix} 5 \\ 5 \\ 4 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 4 \\ 1 \end{pmatrix}$$

El premio Netflix

Sabiendo los puntajes que estos usuarios le dieron a a varias películas podemos calcular los coeficientes de la combinación lineal.

$$a_1 \begin{pmatrix} 1 \\ 3 \\ 5 \\ 2 \\ 3 \end{pmatrix} + a_2 \begin{pmatrix} 2 \\ 3 \\ 5 \\ 4 \\ 5 \end{pmatrix} + a_3 \begin{pmatrix} 3 \\ 1 \\ 4 \\ 5 \\ 2 \end{pmatrix} + a_4 \begin{pmatrix} 1 \\ 4 \\ 5 \\ 1 \\ 4 \end{pmatrix} + a_5 \begin{pmatrix} 5 \\ 5 \\ 4 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 4 \\ 1 \end{pmatrix}$$

Estimando los valores de a_1, a_2, \dots, a_5 , Netflix puede predecir el puntaje que le voy a dar una película nueva si sabe los puntajes que estos 5 usuarios le dieron.

El premio Netflix

Problema: ¿cuántos perfiles usamos para describir a todos los usuarios como combinaciones lineales?

El premio Netflix

Problema: ¿cuántos perfiles usamos para describir a todos los usuarios como combinaciones lineales?

Esto equivale a calcular el rango de la matriz de datos. Este rango es en general muy grande, pero podemos tomar menos perfiles si perdemos "poca información".

El premio Netflix

Problema: ¿cuántos perfiles usamos para describir a todos los usuarios como combinaciones lineales?

Esto equivale a calcular el rango de la matriz de datos. Este rango es en general muy grande, pero podemos tomar menos perfiles si perdemos "poca información".

→ descomposición en valores singulares (SVD)

El premio Netflix

Propuesta 3: no seleccionamos perfiles y usamos a todos los clientes.

$$v_{1354673} = a_1 v_1 + a_2 v_2 + \cdots + a_{1354672} v_{1354672}$$

Tenemos muchas más incógnitas que ecuaciones (la matriz tiene más columnas que filas).

El premio Netflix

Propuesta 3: no seleccionamos perfiles y usamos a todos los clientes.

$$v_{1354673} = a_1 v_1 + a_2 v_2 + \cdots + a_{1354672} v_{1354672}$$

Tenemos muchas más incógnitas que ecuaciones (la matriz tiene más columnas que filas).

→ hay infinitas soluciones.

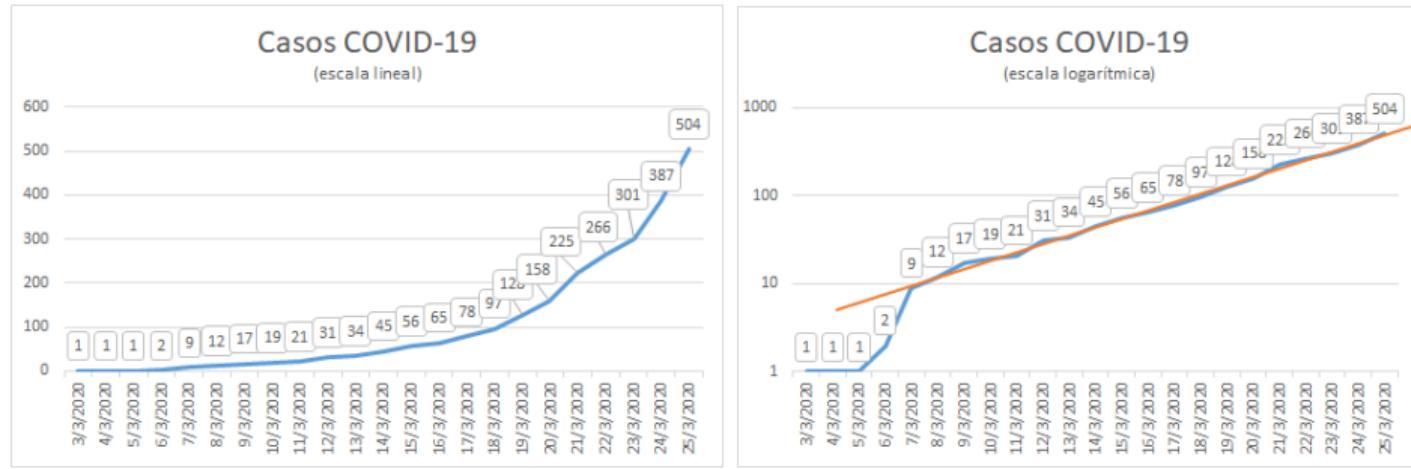
Filas y columnas

En general,

- Más columnas que filas → infinitas soluciones
- Igual cantidad de filas y columnas → solución única
- Más filas que columnas → no hay solución

¿Cuál preferimos?

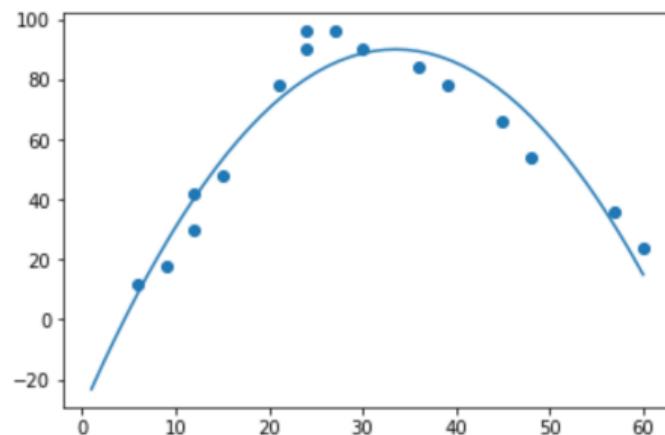
Ejemplo: ajuste de funciones - ¡La vida real!



Fuente: Twitter @jorgeluisaliaga

En la vida real tenemos más filas que columnas. No hay solución, pero buscamos una solución aproximada.

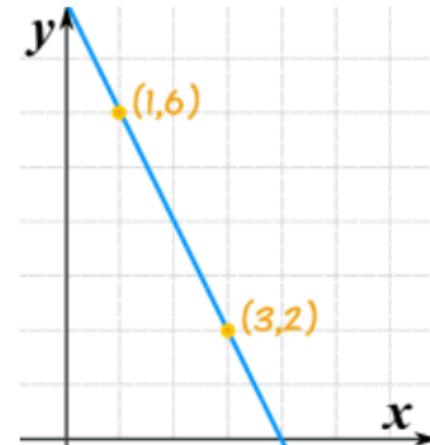
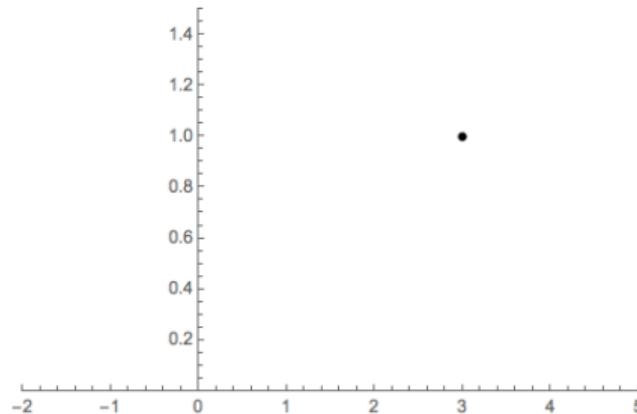
¿Y si queremos ajustar una función cuadrática?



Queremos buscar una función de la forma $f(x) = ax^2 + bx + c$ que aproxime los datos.

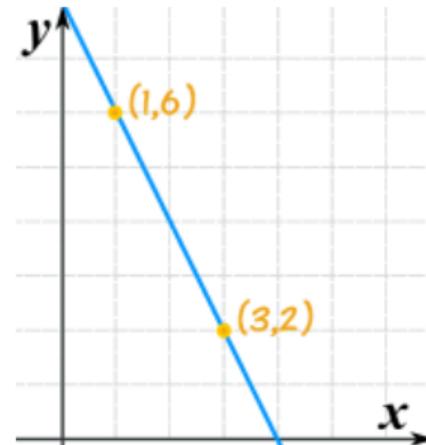
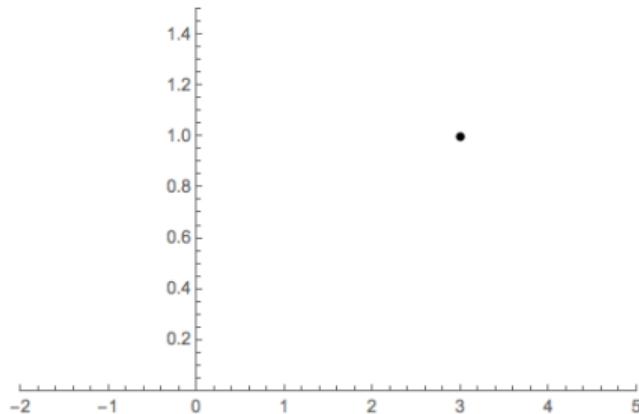
Esto también es un sistema lineal de ecuaciones... ¡aunque usted no lo crea!

¿El sueño de los matemáticos?



- Infinitas soluciones: ¿cómo elegimos una solución?
- Solución única: ¿cómo sabemos si es un buen modelo?

¿El sueño de los matemáticos? La pesadilla de los matemáticos



- Infinitas soluciones: ¿cómo elegimos una solución?
- Solución única: ¿cómo sabemos si es un buen modelo?

Ajústense los cinturones y ...

Ajústense los cinturones y ...

© 2011 Kevin Spear kevin@kevinspear.com www.kevinspear.com



¡largamos!