



MATEMÁTICAS ESENCIALES PARA INTELIGENCIA ARTIFICIAL, APRENDIZAJE AUTOMÁTICO Y CIENCIA DE DATOS

MODELOS LINEALES DE VARIAS VARIABLES

- Un modelo lineal es un modelo que intenta relacionar una variable a partir de una combinación lineal de otras variables. Es decir, usando solamente sumas y multiplicaciones por coeficientes constantes

$$t = ax + by + cz + dw + e$$

- En un modelo lineal Podemos aplicar transformaciones no lineales a las variables de entrada, pero después estas deben combinarse linealmente

$$t = ax + bx^2 + cxy + d\log(z) + ew + \dots$$

MODELOS LINEALES DE VARIAS VARIABLES

- Ejemplo sencillo: DataTemperaturePressureHumidity.xls
- Objetivo: predecir la temperatura a partir de presión y humedad

$$t = a \cdot p + b \cdot h + c$$

- ¿Serás más importante la presión o la humedad? ¿Podemos fijarnos en a y b ?
- → Solo si normalizamos los datos primero.

MODELOS LINEALES DE VARIAS VARIABLES

- Normalización de datos:

- Restar mínimo y dividir por el máximo:

$$\bar{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

(+) TODOS VALORES EN $[0,1]$

(-) MUY SENSIBLE A ERRORES (OUTLIERS)

- Restar media y dividir por desviación típica

$$x = \frac{x - x_{\text{mean}}}{x_{\text{std}}}$$

(+) MÁS ROBUSTO

(-) ASUMIDOS

NORMALES PERO
NO ES ERRÓNEO
EN GENERAL

(-) NO EN
 $[0,1]$

MODELOS LINEALES DE VARIAS VARIABLES

- Ejemplo sencillo: DataTemperaturePressureHumidity.xls

1. Normalizar datos
2. Dividir datos en entrenamiento y validación
3. Ajustar modelos
4. Interpretar resultados
5. Comparar resultados

$$t = a h + b p + c$$

$$t = a \sqrt{h} + b p + c$$

$$t = a h + b$$

$$t = a p + b$$

$$t = ah + b_p + c$$

- Matrices de coeficientes del Sistema de ecuaciones

$$\begin{aligned} a \cdot 0.006992 + b \cdot 1018.41 + c &= 8.3934 \\ a \cdot 0.00673 + b \cdot 1019.81 + c &= 8.62 \end{aligned}$$

$$\begin{bmatrix} h & p & 1 \\ \downarrow & \downarrow & \downarrow \\ 0.006992 & 1018.41 & 1 \\ 0.00673 & 1019.81 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 8.39 \\ 8.62 \\ 9.36 \end{bmatrix}$$

Humidity	Pressure	Temperature
0.006992883	1018.41577	8.39345741
0.006735656	1019.81128	8.62035561
0.006697277	1021.31366	9.36129189
0.006370306	1022.71802	10.4359684
0.00611655	1023.69019	11.2335148
0.005858063	1024.56128	11.6422977
0.005528245	1025.07788	11.7620487
0.005323324	1025.3197	11.5939579
0.005356287	1025.78662	11.1185551
0.005332942	1026.62427	9.62301064
0.004891019	1027.53455	8.70961952
0.004551073	1028.24143	8.16137069

↑
DATOS ENTRENAMIENTO
NORMALIZADOS

MODELOS LINEALES DE VARIAS VARIABLES

- Otro modelo y su matriz de coeficientes del Sistema de ecuaciones:

$$t = a\sqrt{h} + b h + c\sqrt{p} + d p + e$$

$$\begin{bmatrix} \sqrt{h} & h & \sqrt{p} & p & 1 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} t \\ \downarrow \end{bmatrix}$$

MODELOS DINÁMICOS

- Predecir valores futuros de variables a partir de valores pasados
- Conjunto de entrenamiento con valores pasados y test con valores futuros
- Ejemplo: CurrencyExchangeRates.xls
- ¿Cuánta información pasada utilizar para predecir información futura?
- ¿Predecir valores absolutos a partir de valores absolutos o predecir variaciones a partir de variaciones?

¿Con o sin constante?

MODELOS DINÁMICOS

- Podemos necesitar procesar el conjunto de datos si vamos a trabajar con diferencias/incrementos

$$\Delta EUR_{n-1} = EUR_{n-1} - EUR_{n-2}$$

- También para montar el conjunto de datos de entrenamiento dependiendo de la información previa que vayamos a usar

DATO →

EUR_{n-2}	EUR_{n-1}	USD_{n-2}	USD_{n-1}	UK_{n-2}	UK_{n-1}	EUR_n
-------------	-------------	-------------	-------------	------------	------------	---------

DISTINTO FORMATO

DATO →

ΔEUR_{n-2}	ΔEUR_{n-1}	ΔUSD_{n-2}	ΔUSD_{n-1}	ΔUK_{n-2}	ΔUK_{n-1}	ΔEUR_n
--------------------	--------------------	--------------------	--------------------	-------------------	-------------------	----------------

MODELOS DINÁMICOS

- Ejercicio: Utiliza el conjunto CurrencyExchangeRates.xls
 - Ajusta un modelo lineal para predecir los valores de EUR en un día a partir de los valores del día anterior de EUR, US Dollar, UK Pound, Yuan y CAD Dollar
 - Ajusta un modelo lineal que utilice información de los dos días previos
 - Ajusta un modelo lineal para predecir el incremento de EUR en un día a partir de los incrementos del día anterior de EUR, US Dollar, UK Pound, Yuan y CAD Dollar
 - Ajusta un modelo lineal que utilice información de los dos días previos
 - Realiza tests estadísticos de sobreentrenamiento para cada modelo
 - Realiza tests estadísticos para comparar el rendimiento de los modelos