APRENDIZAJE AUTOMÁTICO

Grado en Ingeniería Informática

Grupos 2 y 3

Práctica 1 Ejercicio 2

DECSAI

UNIVERSIDAD DE GRANADA

ÍNDICE

MODUS OPERANDI

1. Cada apartado en un script
2. Los resultados numéricos y/o gráficas han de mostrarse por pantalla
3. El path de lectura de datos de un fichero auxiliar debe ser ".....datos/nombre_fichero"
4. El código ha de ejecutarse completo sin errores
5. Comentar el código es obligatorio
6. Los puntos de parada para mostrar resultados y/o gráficas son necesarios
7. Todos los ficheros se entregan dentro de un zip
8. No entregar datos
9. Subir el zip a PRADO

EJERCICIO 1. REGRESIÓN LINEAL

¿Qué es?

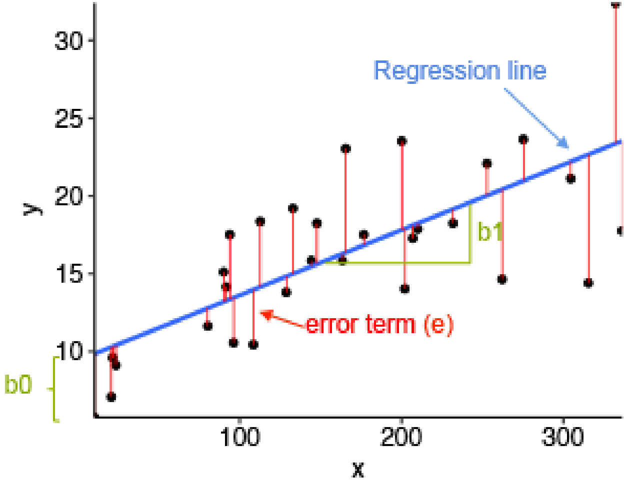
Se trata de un algoritmo matemático que cuantifica el grado de **dependencia** entre una **variable dependiente** (Y) y una o varias **independientes / predictoras** ($X_1, X_2, ...X_n$) mediante una **recta** (en $\mathbb{R}^n$).

Buscamos entonces, determinar las constantes ($w_i \quad con$ $i = 1, 2, ..., n, \quad n \in \mathbb{N}$) de la siguiente ecuación:

$$y = w_0 + w_1 * x_1 + w_2 * x_2 + ...$$

EJERCICIO 1. REGRESIÓN LINEAL

¿Qué es?



EJERCICIO 1. REGRESIÓN LINEAL

Métodos para el cálculo de constantes

1. Gradiente Descendente Estocástico

Es una **variante** del algoritmo Gradiente Descendente que busca disminuir el **costo computacional** y lo hace estudiando sólo una **muestra** del dataset (ejecutando varias veces con subconjuntos de datos aleatorios).

EJERCICIO 1. REGRESIÓN LINEAL
Métodos para el cálculo de constantes

1. Gradiente Descendente Estocástico

- Gradiente desdendente

$$\frac{\partial E_{in}(\mathbf{w})}{\partial w_j} = \frac{\partial}{\partial w_j}\frac{1}{N}\sum_{n=1}^{N}(\mathbf{w}^{\mathrm{T}}\mathbf{x}_{\mathrm{n}} - y_n)^2 = \frac{2}{N}\sum_{n=1}^{N} x_{nj}(\mathbf{w}^{\mathrm{T}}\mathbf{x}_{\mathrm{n}} - y_n) = \boxed{\frac{2}{N}\sum_{n=1}^{N}} x_{nj}(\mathbf{h}(\mathbf{x}_{\mathbf{n}}) - y_n)$$

- Gradiente desdendente (M<N)

$$\frac{\partial E_{in}(\mathbf{w})}{\partial w_j} = \boxed{\frac{2}{M}\sum_{n=1}^{M}} x_{nj}(\mathbf{h}(\mathbf{x}_{\mathbf{n}}) - y_n)$$

EJERCICIO 1. REGRESIÓN LINEAL
Métodos para el cálculo de constantes

1. Gradiente Descendente Estocástico

Ventajas:

- Computación rápida
- Mayor variabilidad en la estimación del gradiente
- Evidencia empírica de generar un buen mínimo local en funciones no convexas
- Muy útil para conjuntos con un gran número de datos

Inconvenientes:

- Puede que no se alcance el óptimo (exacto), pero se halle una aproximación muy razonable

EJERCICIO 1. REGRESIÓN LINEAL

Métodos para el cálculo de constantes

2. Pseudoinversa

La pseudoinversa *R* de una matriz dada *M* es una variante de la inversa, que cumple que al multiplicarla por ambos lados (derecho e izquierdo) genera la identidad.

$$R = A^t (AA^t)^{-1}$$

¿Puede calcularse siempre?

EJERCICIO 1. REGRESIÓN LINEAL
Métodos para el cálculo de constantes

2. Pseudoinversa

Buscamos entonces $y = w_0 + w_1 X$, con w_0, w_1 constantes, y la variable dependiente y X la matriz de las variables independientes.

Minimizing E_{in} : normal equations

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\|\mathrm{X}\mathbf{w} - \mathbf{y}\|^2$$

$$\nabla E_{\text{in}}(\mathbf{w}) = \frac{2}{N}\mathrm{X}^{\intercal}(\mathrm{X}\mathbf{w} - \mathbf{y}) = \mathbf{0}$$

$$\mathrm{X}^{\intercal}\mathrm{X}\mathbf{w} = \mathrm{X}^{\intercal}\mathbf{y}$$

$$\mathbf{w} = \mathrm{X}^{\dagger}\mathbf{y} \quad \text{where} \quad \mathrm{X}^{\dagger} = (\mathrm{X}^{\intercal}\mathrm{X})^{-1}\mathrm{X}^{\intercal}$$

$\mathrm{X}^{\dagger}$ is the '**pseudo-inverse**' of X