Modelos de Aprendizaje

Regresión Lineal Simple

Machine Learning - Luis Zúñiga

Objetivos

1. Explorar nuestro primer algoritmo de aprendizaje supervisado.
2. Recordar la notación.
3. Establecer las ideas principales que rigen los modelos de aprendizaje supervisado.

Regresión Lineal Simple

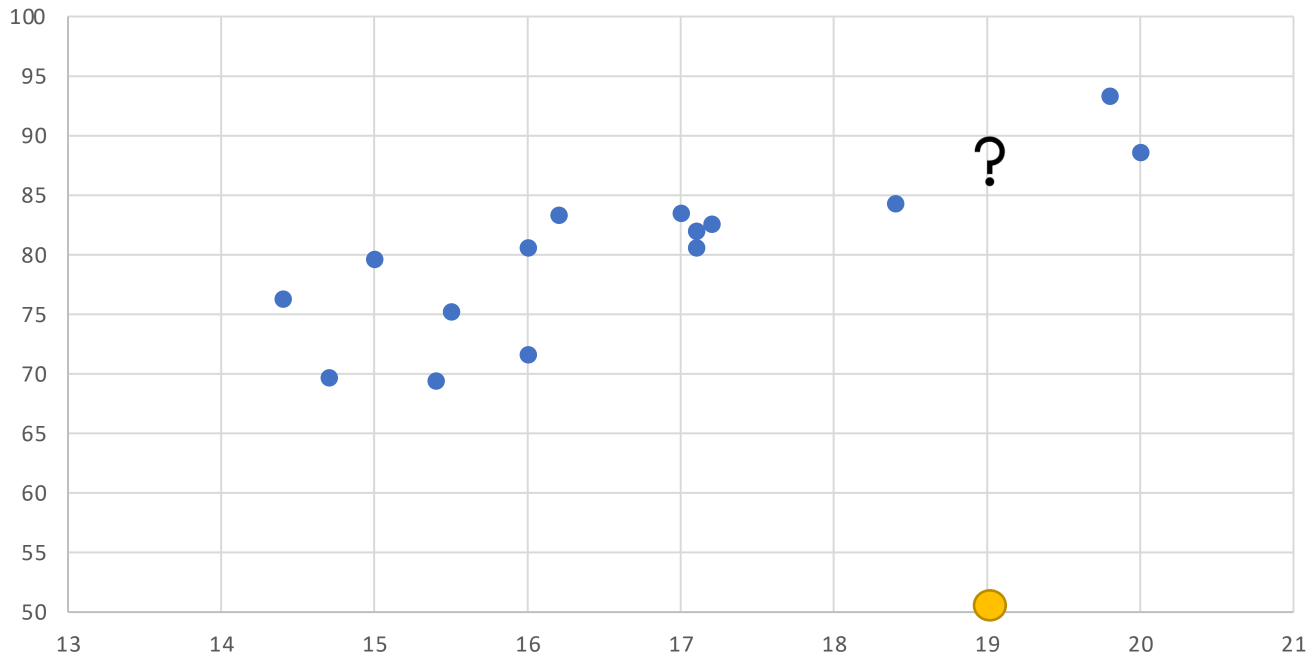
15 datos

x	y
20	88.5999985
16	71.5999985
19.7999992	93.3000031
18.3999996	84.3000031
17.1000004	80.5999985
15.5	75.199997
14.6999998	69.699997
17.1000004	82
15.3999996	69.4000015
16.2000008	83.3000031
15	79.5999985
17.2000008	82.5999985
16	80.5999985
17	83.5
14.3999996	76.3000031

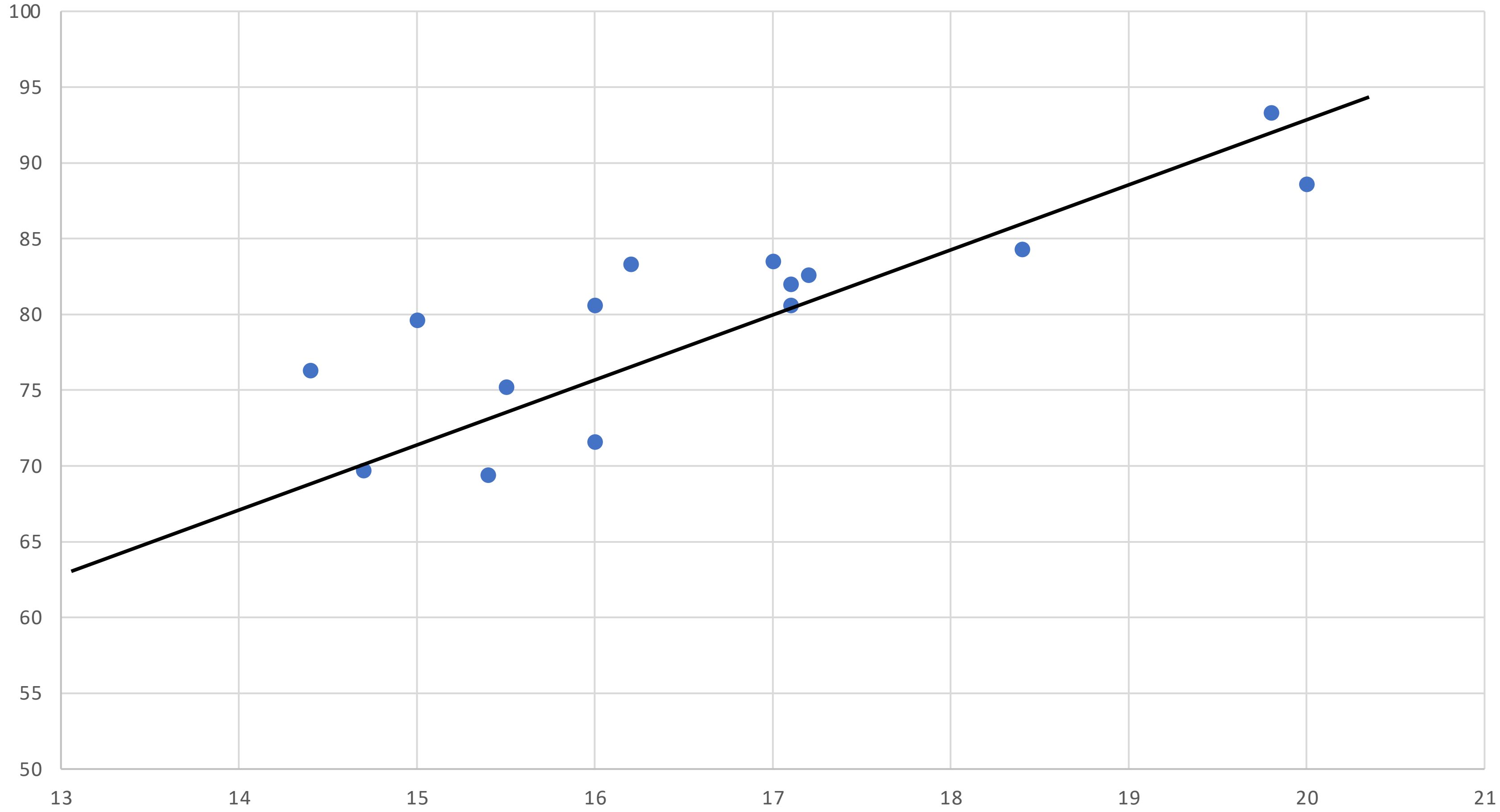
Fuente: [Cengage]

Conjunto de datos

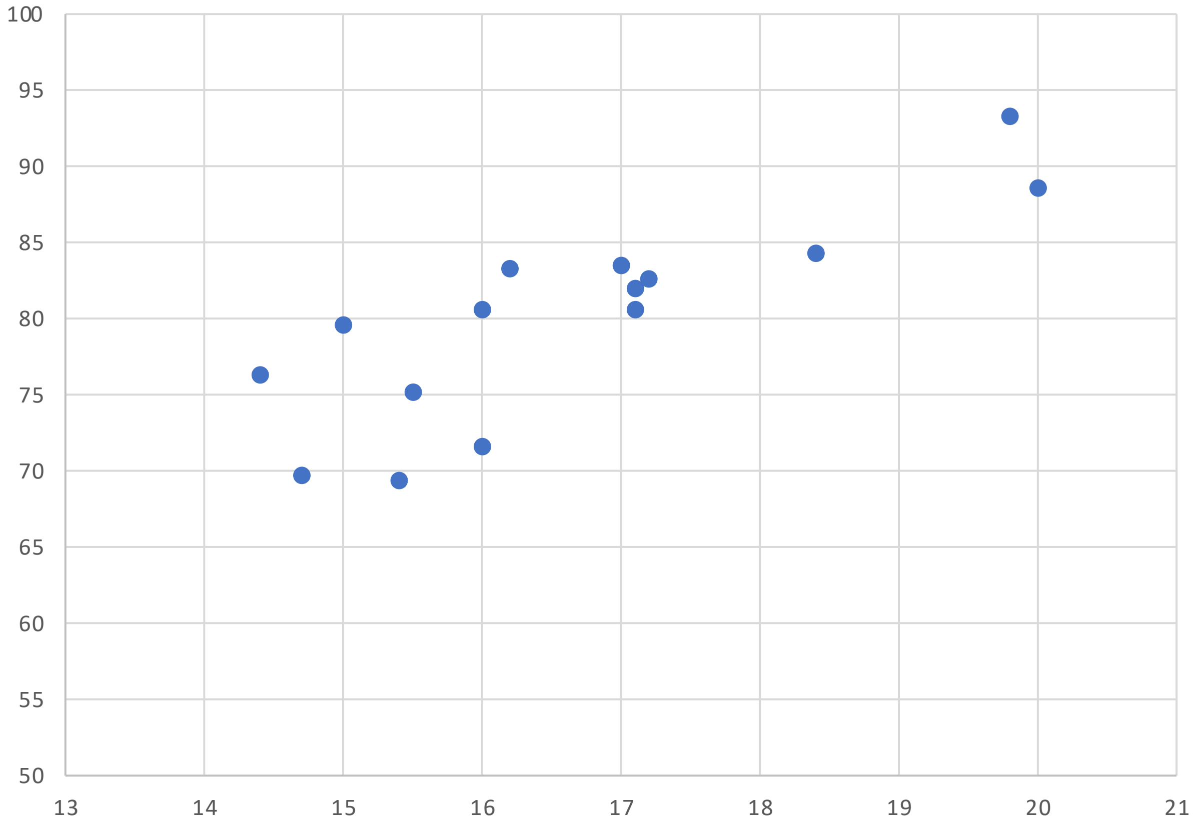
x = chirridos/segundo de un grillo → variable de entrada / características
y = Temperatura en Farenheit → variable de salida / objetivo



Regresión Lineal Simple



Regresión Lineal Simple



¿Con qué otras funciones podrían adaptar los datos?

Idea Principal

Conjunto de entrenamiento
X
$(x_1, y_1), \dots, (x_n, y_n)$

↓

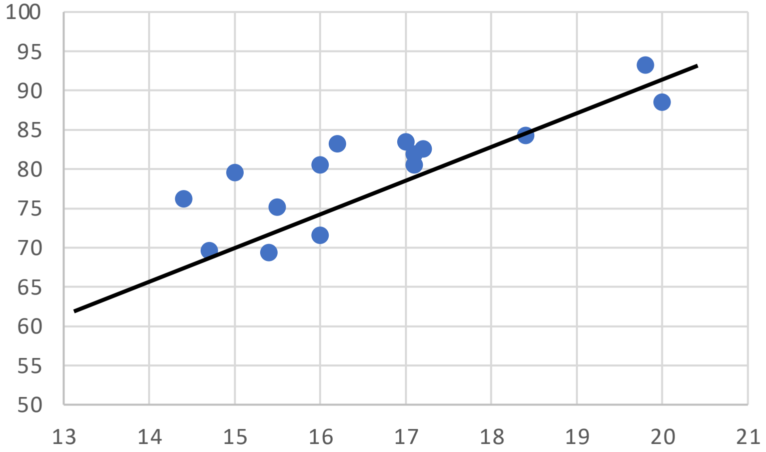
Algoritmo de Aprendizaje

↓

Nuevo x_i → $h: X \rightarrow Y$ → Temperatura y_i

¿Cómo determinar h?

$$h_\theta(x) = \theta_0 + \theta_1 x$$



Regresión Lineal con una variable

Formulación del Problema

Suponiendo que modelamos por medio de una recta:

- Se pueden trazar infinitas opciones.
- ¿Cuál es la mejor?
- ¿Cómo definimos la mejor?
- ¿Cómo se determina?

Regresión Lineal – Estimación de Parámetros

x	y
20	88.5999985
16	71.5999985
19.7999992	93.3000031
18.3999996	84.3000031
17.1000004	80.5999985
15.5	75.199997
14.6999998	69.699997
17.1000004	82
15.3999996	69.4000015
16.2000008	83.3000031
15	79.5999985
17.2000008	82.5999985
16	80.5999985
17	83.5
14.3999996	76.3000031

$$h_{\boldsymbol{\theta}}(x) = \underline{\theta_0} + \underline{\theta_1} x$$

Parámetros del modelo

<u>Problema</u>: Determinar los mejores valores de θ_0 y θ_1.

Dependiendo de sus valores, obtenemos diferentes funciones.

Regresión Lineal – Estimación de Parámetros

Actividad

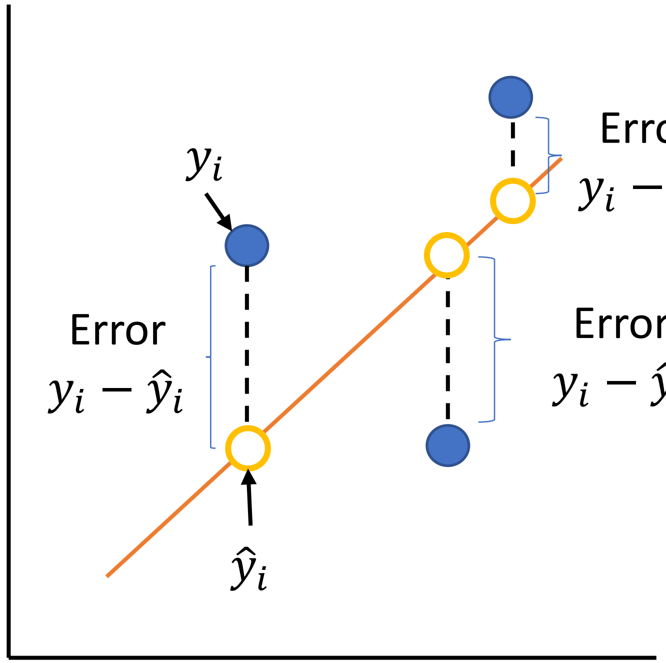
Cada uno:

- Graficar $h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 x$ con distintos parámetros.
- Dibujar la gráfica en el pizarrón.

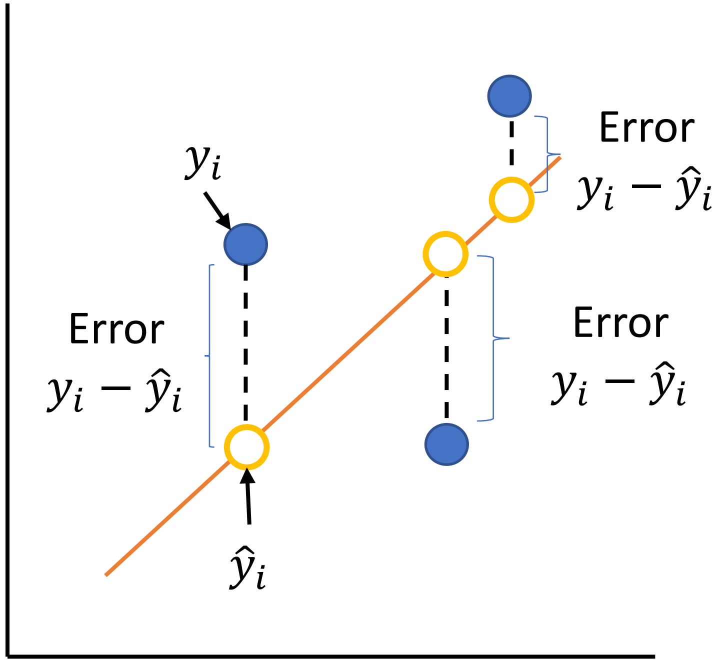
Regresión Lineal – Estimación de Parámetros

Idea: Elegir θ_0 y θ_1 tales que $h(x)$ sea lo más cercana a y en los datos de entrenamiento (x, y).

Otra forma de interpretarlo: minimizar la diferencia o el error de $h(x)$ respecto a los datos de entrenamiento (x, y).



Regresión Lineal – Estimación de Parámetros



Minimizar el error de todos los puntos:

$$\min_{\theta_0, \theta_1} \frac{1}{2n} \sum_{i=1}^{n} (h_{\boldsymbol{\theta}}(x_i) - y_i)^2$$

Valor que predice el modelo
$\hat{y} = h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 x$

Valor verdadero del conjunto de datos

A esto se le conoce como **función de costo**

Regresión Lineal – Estimación de Parámetros

Hipótesis o modelo:

$$\hat{y} = h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 x$$

Función de costo:

$$\min_{\theta_0, \theta_1} \frac{1}{2n} \sum_{i=1}^{n} (h_{\boldsymbol{\theta}}(x_i) - y_i)^2$$

Para resumir:

- Esta función de costo se le conocer como **error cuadrático medio**.
- Esto es un problema de optimización.
- En particular, se buscan los parámetros θ_0, θ_1 que minimicen la función de costo propuesta.
- ¿Es la única función de costo que se puede usar? No, pero es la más común, sencilla de usar y da buenos resultados.
- *¿Cómo se resuelve esto?*

Regresión Lineal – Estimación de Parámetros

Actividad

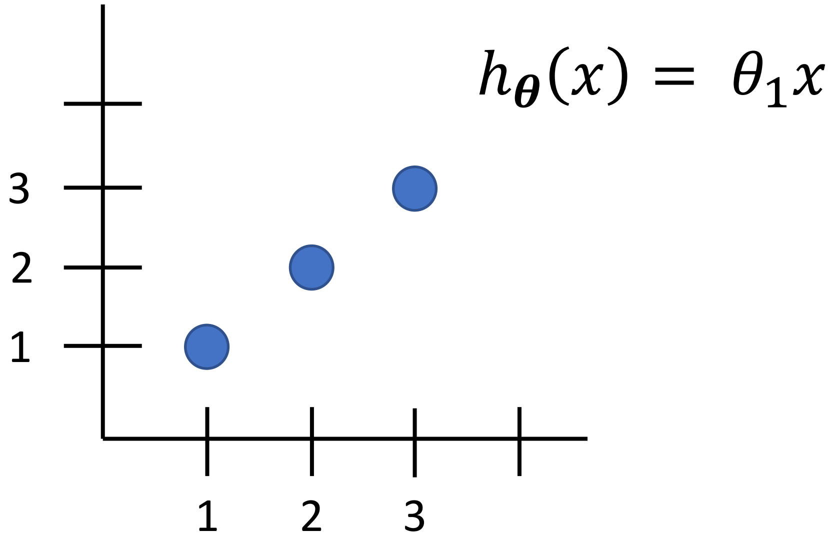
Consideremos un modelo simplificado

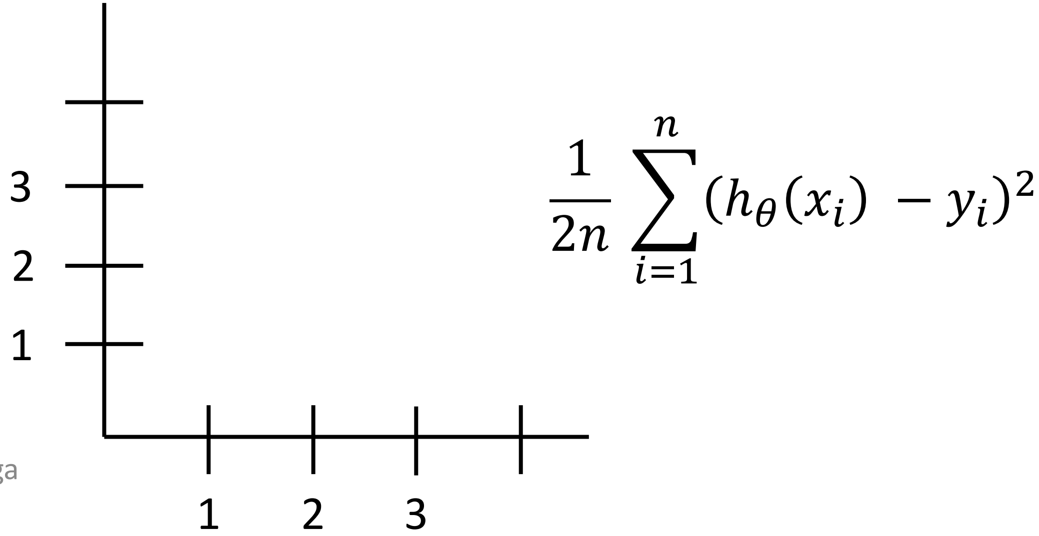
$$h_{\boldsymbol{\theta}}(x) = \theta_1 x$$

para modelar datos, y la función de costo

$$\frac{1}{2n}\sum_{i=1}^{n}(h_{\boldsymbol{\theta}}(x_i) - y_i)^2$$

¿Cómo se vería la función de costo con distintos valores de θ_1?





Regresión Lineal – Estimación de Parámetros

Problema:

$$J = \min_{\theta_0, \theta_1} \frac{1}{2n} \sum_{i=1}^{n} (h_{\boldsymbol{\theta}}(x_i) - y_i)^2$$

Solución:

Determinar

$$\frac{\partial J}{\partial \theta_0} = 0, \frac{\partial J}{\partial \theta_1} = 0$$

Actividad

Determinar:

$$\frac{\partial}{\partial \theta_0} \frac{1}{2n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_i - y_i)^2 = 0$$

$$\frac{\partial}{\partial \theta_1} \frac{1}{2n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_i - y_i)^2 = 0$$

Regresión Lineal – Estimación de Parámetros

Problema:

$$J = \min_{\theta_0, \theta_1} \frac{1}{2n} \sum_{i=1}^{n} (h_{\boldsymbol{\theta}}(x_i) - y_i)^2$$

Solución:

Determinar

$$\frac{\partial J}{\partial \theta_0} = 0, \frac{\partial J}{\partial \theta_1} = 0$$

Solución:

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$\theta_1 = \frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})}$$

donde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Regresión Lineal – Resumen

Hipótesis o modelo:

$$\hat{y} = h_{\boldsymbol{\theta}}(x) = \theta_0 + \theta_1 x$$

Función de costo:

$$\min_{\theta_0, \theta_1} \frac{1}{2n} \sum_{i=1}^{n} (h_{\boldsymbol{\theta}}(x_i) - y_i)^2$$

Soluciones:

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

$$\theta_1 = \frac{\sum x_i (y_i - \bar{y})}{\sum x_i (x_i - \bar{x})}$$

Actividad

Con los datos de los grillos, determinar la ecuación de la recta que mejor se ajuste a los datos.

Regresión Lineal – Estimación de Parámetros

Tarea

Demostrar que

$$\theta_1 = \frac{\sum x_i(y_i - \bar{y})}{\sum x_i(x_i - \bar{x})}$$

es equivalente a

$$\theta_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

<u>Nota</u>:

$$c\sum(z_i - \bar{z}) = 0$$

para cualquier c y z. *(¿Por qué?)*

Estimación de Parámetros

- Este método *analítico* para estimar parámetros es bueno.
- Desafortunadamente, se debe determinar de antemano. Una computadora no puede calcular (*fácilmente*) derivadas.
- Debemos utilizar métodos numéricos para acelerar el proceso de determinar mínimos globales*.
- Esta es una de las diferencias con el enfoque estadístico.

Gradiente Descendiente

Veamos cuál es la idea del gradiente descendiente…

Gradiente Descendiente: ¿Cómo se ve en nuestro problema?

$$\theta_0 = \theta_0 - \eta \left(\frac{\partial}{\partial \theta_0} \frac{1}{2n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_i \ - y_i)^2 \right)$$

$$= \theta_0 - \eta \left(\frac{\sum \theta_0 + \theta_1 x_i \ - y_i}{n} \right)$$

$$\theta_1 = \theta_1 - \eta \left(\frac{\partial}{\partial \theta_1} \frac{1}{2n} \sum_{i=1}^{n} (\theta_0 + \theta_1 x_i \ - y_i)^2 \right)$$

$$= \theta_1 - \eta \left(\frac{\sum (\theta_0 + \theta_1 x_i \ - y_i) x_i}{n} \right)$$

Gradiente Descendiente: ¿Cómo se ve en nuestro problema?

1. Se eligen valores iniciales de θ_0, θ_1 al azar.
2. Repetir hasta convergencia:

$$\theta_0 = \theta_0 - \eta \left(\frac{\sum \theta_0 + \theta_1 x_i \ - y_i}{n} \right)$$

$$\theta_1 = \theta_1 - \eta \left(\frac{\sum (\theta_0 + \theta_1 x_i - y_i) x_i}{n} \right)$$

Batch Gradient Descent

- Esta forma de Gradiente Descendiente se conoce como *Batch*.
- Ya que en cada iteración utiliza todos los puntos del conjunto de datos.
- Existen otras versiones que utilizan solo un subconjunto de los puntos, no todo.

Conclusiones

Conjunto de datos

Entrenamiento del Algoritmo

Función de Costo

Optimización (minimizar el error o costo)

- Vieja confiable de cálculo
- *(Batch) Gradient Descent*

Modelo que «Aprendió»

Para finalizar…

¡Felicidades! Acaban de aprender su primer método de aprendizaje supervisado.

- Un algoritmo clásico de la estadística desde la perspectiva del ML.
- Un método numérico para optimizar.
- La idea detrás del flujo de los métodos de aprendizaje supervisado.

Fin de la presentación

¡Gracias por su preferencia!