Regresión lineal

Dra. Ximena Gutierrez Vasques & Mtro. Víctor Mijangos

Facultad de Ingeniería

21 de enero, 2020

Regresión lineal

Un ejemplo clásico del aprendizaje es la **regresión lineal** [1].
La idea general es encontrar una **función lineal** que describa el comportamiento de los datos.
Es decir, dado unas variables de entrada $X_1, X_2, \ldots, X_d$, queremos predecir un valor de salida Y.
Por lo que debemos encontrar:

$$f(x) = y$$

tal que $f : \mathbb{R}^d \to \mathbb{R}$ y f lineal.

Regresión lineal

Una función lineal es aquella que cumple:

$$f(\lambda x + x') = \lambda f(x) + f(x')$$

Y puede verse que las funciones lineales de $\mathbb{R}^d$ a $\mathbb{R}$ son de la forma:

$$f(x) = wx$$

Tal que $w \in (\mathbb{R}^d)^*$.
En la práctica se utiliza un parámetro de sesgo o bias, tal que las funciones son:

$$f(x) = wx + b \tag{1}$$

con $w \in (\mathbb{R}^d)^*$, $b \in \mathbb{R}$.

Regresión lineal

Ejemplo

Supóngase que se tiene una lista de casas, tal que se conoce el número promedio de cuarto (X) y el precio de la casa (Y):

	X	**Y**
Casa 1	6	22.9
Casa 2	7.14	36.2
Casa 3	6.4	21.6
Casa 4	6.7	30.5
Casa 5	5.1	16.3
Casa 6	7.15	37.3
Casa 7	8.3	50
Caa 8	8.2	48

Regresión lineal

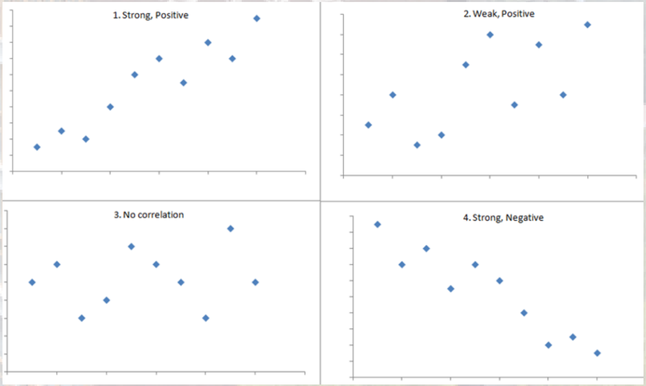
Pueden plantearse la preguntas:

- ¿Existe una correlación (lineal) entre las variables X e Y.
- Si existe ¿cómo se comporta esta correlación?

A partir de la regresión lineal, podemos estimar la correlación que existe entre una y otra variable.

Regresión lineal y correlación lineal

Una correlación estima la dependencia entre dos variables aleatorias. Una alta correlacion asegura que la regresión lineal se acople a los datos adecuadamente.



Residuos en regresión lineal

El objetivo es entonces predecir el valor de salida y de un valor de entrada x. Podemos estimar que el valor y es [1]:

$$y = wx + b + \epsilon_y$$

Aquí, ϵ_y es un residuo.

Residuos en regresión lineal

Despejando la función anterior, obtenemos que los residuos se calculan como:

$$\epsilon_y = y - wx + b$$

O bien, como:

$$\epsilon_y = y - f(x)$$

Función de riesgo en regresión

Para garantizar que la regresión es adecuada, se busca que la suma de los residuos sea pequeña. Surgen dos aproximaciones:

- **Least-absolute value (LAV)**:

$$R(f) = \sum_{\mathcal{S}} |y - f(x)|$$

- **Least-squares**:

$$R(f) = \sum_{\mathcal{S}} (y - f(x))^2$$

El método de least-squares representa mayor sencillez para resolverse.

Regresión lineal como problema de aprendizaje

Hasta ahora contamos con los siguientes elementos:

1. Un conjunto de datos supervisados $\mathcal{S} = \{(x, y) : x \in \mathbb{R}^d, y \in \mathbb{R}\}$ (se espera que exista una correlación lineal entre X y Y).
2. Una función que define la ML:
$$f(x) = wx + b$$
tal que $w \in \mathbb{R}^d$ y $b \in \mathbb{R}$ son los parámetros a aprender.
3. Una función de riesgo:
$$R(f) = \frac{1}{2}\sum_{\mathcal{S}}(y - f(x))^2$$
(El factor $\frac{1}{2}$ ayuda a simplificar la derivación).

Regresión lineal como problema de aprendizaje

El objetivo es encontrar una función $\hat{f}$ (dependiente de w y b) que minimice la función de riesgo.
Asumimos que R es convexa, entonces buscamos su punto de inflexión tal que:

$$\nabla_w R(f) = 0$$

De aquí que:

$$\nabla_w \frac{1}{2} \sum_{\mathcal{S}} (y - f(x))^2 = 0$$

$$\nabla_w \frac{1}{2} ||Y - Xw||^2 = 0$$

$$X^T Y - X^T X w = 0$$

$$(X^T X)^{-1} X^T Y = w$$

Regresión lineal como problema de aprendizaje

Se ha incorporado el bias al vector w ($[w; b]$). Y es el vector de valores esperados y X la matriz cuyos renglones son los ejemplos.
Tenemos, entonces que:

$$\arg\min_{w} \frac{1}{2}||Y - Xw||^2 = (X^T X)^{-1} X^T Y$$

Regresión como modelo paramétrico

La regresión lineal es un **modelo paramétrico**, en tanto asume propiedades de la distribución de los datos.

Desde una perspectiva probabilística, la regresión lineal busca estimar $\mathbf{p}(\mathbf{y})$ asumiendo que [1]:

$$y \sim N(f(x), 1)$$

Es decir, asume una distribución normal.

Distribución normal

En este sentido, la función de probabilidad depende de dos parámetros: **media** y **varianza**. Esta última se asume igual a 1, tal que:

$$p(y) = \frac{1}{\sqrt{2\pi}} exp[-\frac{1}{2}(y - f(x))^2]$$

El objetivo entonces es estimar la media $f(x)$ con respecto a los datos X de entrada.

Relación con la entropía

Dada la función de probabilidad (que depende de f, ergo de w) se busca minimizar la entropía:

$$R(p) = -\sum_{y} \ln p(y)$$

De aquí:

$$\begin{aligned}
\arg\min R(p) &= \arg\min -\sum_{y} \ln \frac{1}{\sqrt{2\pi}} exp[-\frac{1}{2}(y - f(x))^2] \\
&= \arg\min -\sum_{y} [\ln \frac{1}{\sqrt{2\pi}}] - \frac{1}{2}(y - f(x))^2 \\
&= \arg\min \frac{1}{2} \sum_{\mathcal{S}} (y - f(x))^2
\end{aligned}$$

References

John Fox.
Applied regression analysis and generalized linear models.
Sage Publications, 2015.

The End