Apunte clase de Inteligencia Artificial - Regresión lineal

César Jiménez Salazar

22 de Febrero del 2024

1 Regresión lineal

El análisis de regresión lineal se utiliza para predecir el valor de una variable (y) en función del valor de otra variable (x). La variable que desea predecir se llama variable dependiente (x). La variable que se utiliza para predecir el valor de la otra variable se llama variable independiente (y). [1]

En el contexto de regresión lineal para la inteligencia artificial se tiene lo siguiente:

- Se tiene una colección de ejemplos etiquetados $\{(x_i, y_i)\}^N{}_{i=1}$
- Donde N es el tamaÑo de la colección.
- x_i Es un vector D-dimensional, donde cada feature $x_i^{(j)}, j = 1, \ldots, D$
- y_i Es la etiqueta, y es un valor que pertenece al conjunto de los números reales.

¿Con esto qué se quiere lograr?

- Tenemos una combinación lineal con los features
- El modelo está parametrizado con w y b
- Encontrar los valores óptimos de w y b que hagan que mi función realice las predicciones más acertadas.
- Es importante recordar que óptimo **no significa** perfecto

En palabras más sencillas, se quiere hacer que el modelo pase lo más cerca posible por cada uno de los puntos, como se puede observar en la figura 1

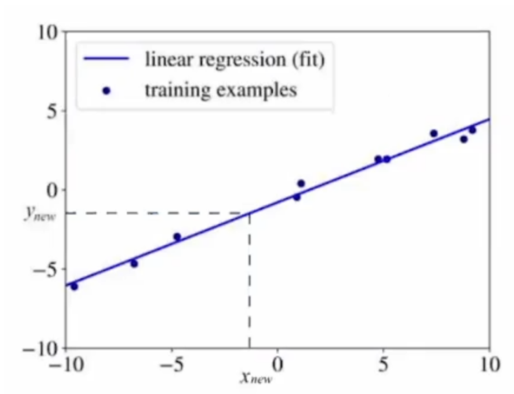

Figure 1: Regresión de una dimensión

2 Loss function

El loss function (función de pérdida) es una medida que cuantifica la discrepancia entre las predicciones de un modelo y los valores reales del conjunto de datos. Para la regresión lineal, que es un tipo de modelo utilizado para predecir valores numéricos, el objetivo es minimizar esta función de pérdida. Se representa de la siguiente forma:

MSE

$$L = \frac{1}{N} \sum (f_{w,b}(x_i) - y_i)^2, i = 1, \ldots, N \tag{1}$$

MAE

$$L = \frac{1}{N} \sum \mid f_{w,b}(x_i) - y_i \mid, i = 1, \ldots, N \tag{2}$$

¿Por qué escoger MSE y no MAE?

1. MSE es derivable, MAE no lo es
2. Refleja mejor la severidad de los errores

2.1 Derivada parcial del loss function

Con respecto a w

$$L = \frac{1}{N} \sum ((wx_i + b) - y_i)^2 \tag{3}$$

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum(((wx_i + b) - y_i)^2)' \tag{4}$$

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot ((wx_i + b) - y_i)' \tag{5}$$

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot ((wx_i + b)' - (y_i)') \tag{6}$$

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot ((x_i)) \tag{7}$$

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot x_i \tag{8}$$

Con respecto a b

$$L = \frac{1}{N}\sum((wx_i + b) - y_i)^2 \tag{9}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum(((wx_i + b) - y_i)^2)' \tag{10}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot ((wx_i + b) - y_i)' \tag{11}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot ((wx_i + b)' - (y_i)') \tag{12}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot ((0 + 1) - 0) \tag{13}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \cdot 1 \tag{14}$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum 2((wx_i + b) - y_i) \tag{15}$$

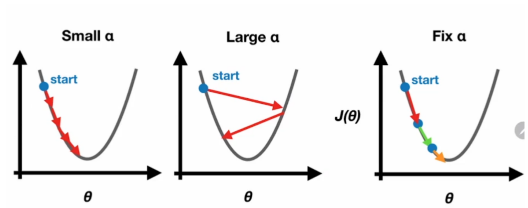

Figure 2: Diferentes learning rates

2.2 Actualización de parámetros

$$w = w - \alpha \frac{\partial L}{\partial w} \tag{16}$$

$$b = b - \alpha \frac{\partial L}{\partial b} \tag{17}$$

1. α es un hiperparámetro (Learning rate)
2. Gradiente descendiente

En la figura se puede observar lo que diferentes valores para α logran en la práctica para encontrar el mínimo local de la función.

Epoch: La cantidad de veces que recorremos el set de datos de entrenamiento para ajustar nuestro modelo.

3 Referencias bibliográficas

[1] "What is linear regression? — IBM." https://www.ibm.com/topics/linear-regression