Regresión

Jose Luis Paniagua Jaramillo
jlpaniagua@uao.edu.co

Agenda

Agenda

Agenda

Regresion

- La regression es un tipo de aprendizaje supervisado, donde el objetivo es predecir resultados con valores continuos.
- A partir de una serie de variables llamadas **predictoras/exploratorias** y una variable de respuesta **continua** llamada **resultado/objetivo**, se trata de encontrar una relación entre dichas variables que permita predecir un resultado.

Ejemplo

Supongamos que estamos interesados en predecir la nota de los estudiantes del curso en el examen 1. Si existe una relación entre el tiempo dedicado a estudiar para el examen y la nota obtenida, podríamos usarla como datos de entrenamiento para aprender un modelo que usa el tiempo de estudio para predecir las notas de los exámenes de los futuros estudiantes que planean matricular este curso.

Agenda

Agenda

Modelo de Regresión Lineal I

Forma Explicita

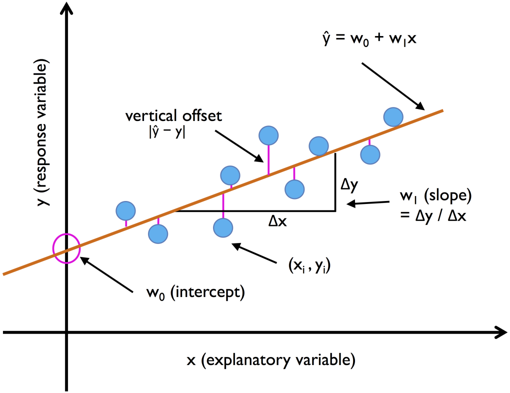


Figura: Regresión Lineal Simple[1]

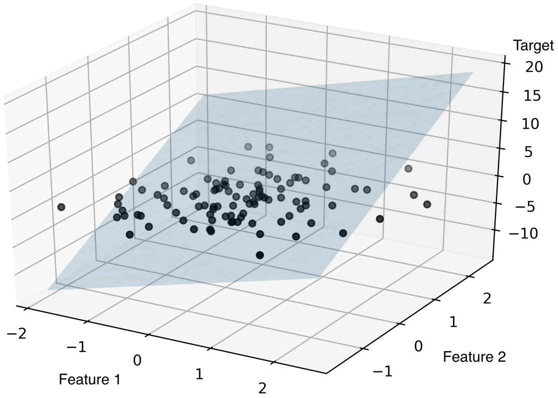


Figura: Regresión Lineal Multiple[1]

Modelo de Regresión Lineal III

Forma Explicita

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

donde:

- $\hat{y}$ es la predicción.
- n es el numero de características.
- x_i es el $i - esimo$ valor de la característica.
- θ_j es el $j - esimo$ parámetro del modelo.
- θ_0 es el **bias** del modelo (intercepto).

Agenda

Modelo de Regresión Lineal I

Forma Matricial

$$\hat{y} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta} \cdot \mathbf{x}$$

donde:

- $\boldsymbol{\theta}$ es el vector de parámetros del modelo.
- $\mathbf{x}$ es el vector de instancias de las características.
- $\boldsymbol{\theta} \cdot \mathbf{x}$ es el producto punto de los vectores $\boldsymbol{\theta}$ y $\mathbf{x}$.
- $h_{\boldsymbol{\theta}}$ es la función de hipótesis.

Nota

- x_0 es siempre igual a 1.
- $\boldsymbol{\theta}$ contiene el bias θ_0.

Agenda

Agenda

Función de Costo

Mean Square Error

$$MSE(\boldsymbol{X}, h_{\boldsymbol{\theta}}) = \frac{1}{m}\sum_{i=1}^{m}(\boldsymbol{\theta}^T \mathbf{x}^{(i)} - y^{(i)})^2$$

El objetivo es encontrar $\boldsymbol{\theta}$ que minimice la función de costo.

Agenda

Regularización

- **Overfitting** es un problema comun en ML, en donde un modelo funciona bien con los datos de entrenamiento, pero no es capaz de funcionar bien ante datos nunca vistos (datos de validacion).
- **underfitting** es un problema opuesto al overfitting, en donde un modelo no funciona bien con los datos de entrenamiento.

Cual es la solucion?

Regularizacion

$$J(\theta) = MSE(\theta) + \alpha \frac{1}{2} \sum_{i=1}^{n} \theta_i^2$$

Agenda

Ecuación Normal (Mínimos Cuadrados)

La **ecuación normal** es una forma directa de encontrar los valores de $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y}$$

donde:

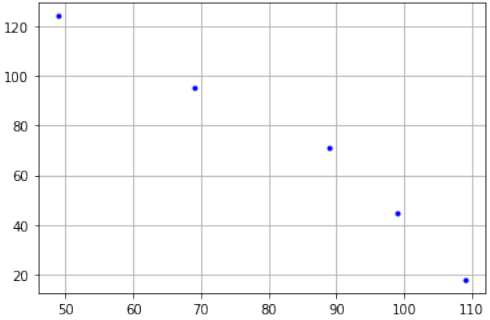
- $\hat{\boldsymbol{\theta}}$ es el valor de $\boldsymbol{\theta}$ que minimiza la función de costo (MSE).
- $\boldsymbol{Y}$ es el vector (matriz) de valores objetivo (variable dependiente) que contiene y_1 hasta y_n
- $\boldsymbol{X}$ es el vector (matriz) de valores de las variables independientes (características) que contiene x_1 hasta x_n.

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

Ecuacion Normal (Mínimos Cuadraros)

Ejmeplo

Price(x)	Demand(y)
49	124
69	95
89	71
99	45
109	18



$$\boldsymbol{Y} = \begin{bmatrix} 124 \\ 95 \\ 71 \\ 45 \\ 18 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & 49 \\ 1 & 69 \\ 1 & 89 \\ 1 & 99 \\ 1 & 109 \end{bmatrix}$$

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{Y} = \frac{1}{11600}\begin{bmatrix} 36765 & -415 \\ -415 & 5 \end{bmatrix} \cdot \begin{bmatrix} 353 \\ 25367 \end{bmatrix} \approx \begin{bmatrix} 211 \\ -1{,}7 \end{bmatrix}$$

$$\hat{y} = 211 - 1{,}7x$$

Agenda

Ejemplo: diabetes dataset

Samples total	442
Dimensionality	10
Features	real, -.2 < x < .2
Targets	integer 25 - 346

Agenda

Gradiente Descendente

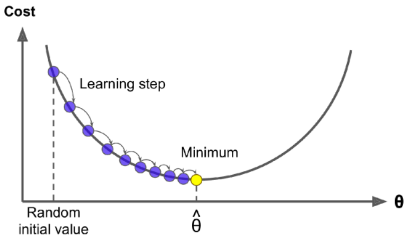


Figura: [2]

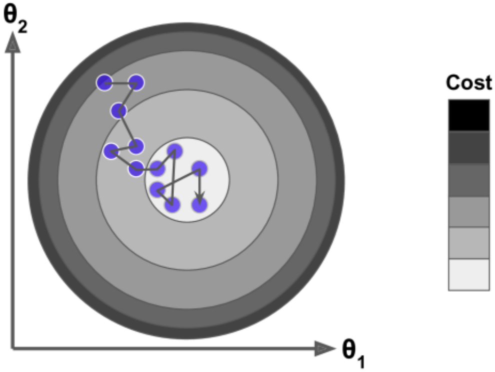
La función de costo MSE para un modelo de Regresión Lineal es una **función convexa**.
Gradiente de la función de consto:

$$\nabla_{\boldsymbol{\theta}} MSE(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial}{\partial \theta_0} MSE(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_1} MSE(\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial}{\partial \theta_n} MSE(\boldsymbol{\theta}) \end{bmatrix} = \frac{2}{m} \mathbf{X}^T (\mathbf{X}\boldsymbol{\theta} - \mathbf{y})$$

Actualización del gradiente:

$$\boldsymbol{\theta}^{(\text{next step})} = \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} MSE(\boldsymbol{\theta})$$

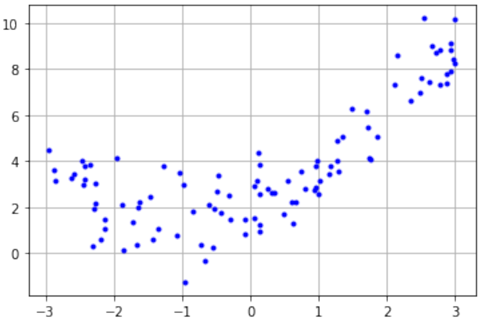
Gradiente Descendente Estocástico

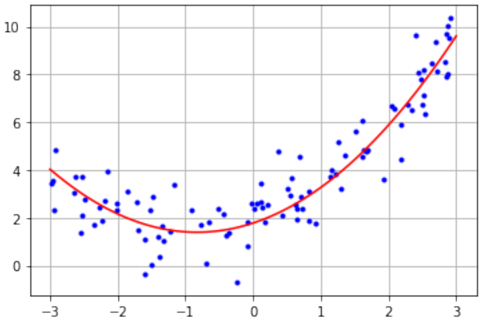


- selecciona de manera aleatoria una parte de los datos de entrenamiento para calcular el gradiente.
- es menos regular que el gradiente convencional.
- permite encontrar valore mínimos de los parámetros, pero no los óptimos.
- debido a su aleatoriedad, tiene mayor probabilidad de encontrar el mínimo global en funciones no convexas.

Agenda

Regresión Polinomial





La técnica consiste en adicionar una nueva variable predictora la cual es el cuadrado de la variable predictora original.

$$\hat{y} = \theta_0 + \theta_1 x_1 \rightarrow \hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_1^2$$

Agenda

Referencias

Sebastian Raschka and Vahid Mirjalili.
Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2.
Packt Publishing Ltd, 2019.

Aurélien Géron.
Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.
O'Reilly Media, 2019.

`https://medium.com/swlh/linear-regression-from-scratch-4ac1cc666ee2`
`https://scikit-learn.org/stable/index.html`