

Curso Complementario R + Python

REGRESIONES

Agenda

Regresión: Conceptos

Regresión Lineal Simple

Regresión Lineal Múltiple

Evaluación de Modelos

Regresión Polinómica

Regresión:Conceptos

Regresión

Regresión: Categoría de Aprendizaje Supervisado para predecir variables en una escala continua.



Aprendizaje
Supervisado

- **Clasificación:** Predicción de variables categóricas (discretas).
- **Regresión:** Predicción de valores de variables continuas.

Regresión Lineal

El objetivo de la Regresión Lineal es modelar la relación entre una o varias características (features) y una variable de destino (target) continua.

Hay dos tipos de regresiones lineales:

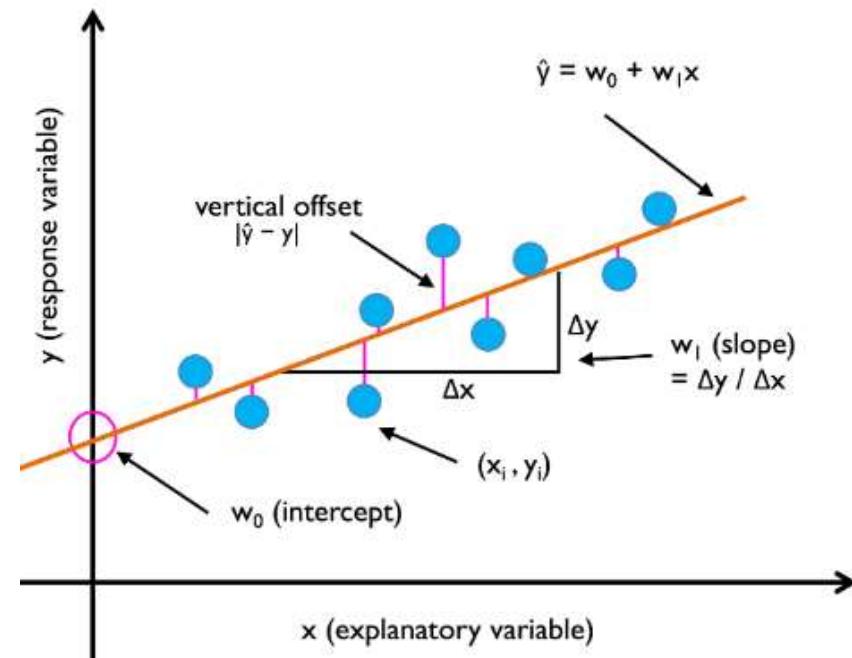
- Regresión Lineal Simple
- Regresión Lineal Múltiple

Regresión Lineal Simple

La Regresión lineal simple (univariada) intenta modelar la relación entre una característica (variable explicativa x) y una respuesta de valor continuo (variable destino o target y)

La ecuación del modelo de Regresión Lineal Simple es simplemente una recta:

$$y = w_0 + w_1 x$$



Regresión Lineal Múltiple

La Regresión Lineal Múltiple es una generalización de la Regresión Lineal Simple, donde consideramos varias variables explicativas.

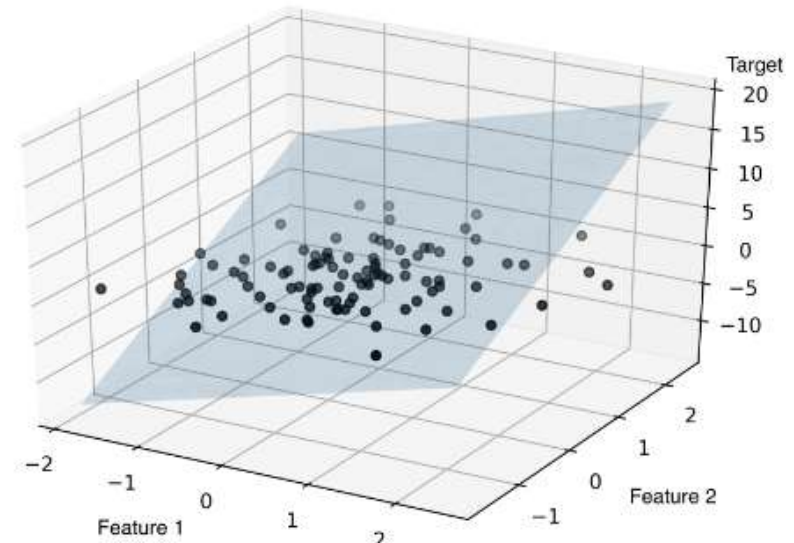
Se modela mediante la siguiente ecuación:

$$y = w_0x_0 + w_1x_1 + \dots + w_mx_m$$

En forma vectorial:

$$y = w^T x$$

(La intersección con eje y es w_0 ; $x_0 = 1$)



Análisis para Regresión

Matriz de Dispersión: para visualizar las relaciones entre las distintas features y ver si hay valores atípicos.

Matriz de Correlación: para cuantificar y resumir las relaciones lineales entre las variables.

(ver notebook rel lineal entre RM y MEDV:

<https://colab.research.google.com/drive/18ZhheYRTVlvcuKymd7qBjEDmw8RhJIhN?usp=sharing>

)

Análisis para Regresión

EN la Regresión Lineal lo que buscamos es la recta de mejor ajuste a través de los puntos de entrenamiento.

Estimación Coef. Regresión Lineal

SKLEARN

Estimación de Coeficientes: SKLEARN

Librería LIBLINEAR, función LinearRegression: me da la pendiente y la ordenada de la recta de regresión.

```
From sklearn.linear_model import LinearRegression
```

Ver notebook

<https://colab.research.google.com/drive/18ZhheYRTVlvcuKymd7qBjEDmw8RhJlhN?usp=sharing>

Ajuste de Modelos con Outliers

RANSAC (RANdom Sampler Consensus) : método robusto de regresión lineal

Proceso de Algoritmo:

1. Elige un numero aleatorio de muestras para ser inliers y ajusta el modelo.
2. Prueba los demás puntos de datos frente al modelo entrenado y agrega aquellos que caen dentro de una tolerancia especificada.
3. Entrena nuevamente el modelo con los inliers
4. Estima el error del modelo ajustado frente a los inliers.
5. Si se alcanza el numero de iteraciones especificada o se cumple la condición de umbral, el algoritmo finaliza, sino vuelve al inicio.

Ajuste de Modelos con Outliers

RANSAC (RANdom Sampler Consensus) : método robusto de regresión lineal

Librería LIBLINEAR, clase RANSACRegressor selecciona un número aleatorio de muestras par ser inliers y ajusta el modelo.

```
From sklearn.linear_model import RANSACRegressor
```

Evaluación de Rendimiento

Error Cuadrático Medio (MSE): es el valor medio del SSE.

Coeficiente de Determinación (r^2 – Rscore)

Gráfico de Residuos

Desviación Absoluta de la Mediana (MAD)

Evaluación de Rendimiento

Error Cuadrático Medio (MSE): es el valor medio del SSE.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

(SSE - Error Suma de cuadrados : es la suma de las diferencias al cuadrado entre cada observación y la media de su grupo.)

Coefficiente de Determinación (r^2 – Rscore)

$$R^2 = 1 - \frac{SSE}{SST}$$

Evaluación de Rendimiento

Desviación Absoluta Mediana (MAD) es una medida robusta de cuán disperso está un conjunto de datos. La varianza y la desviación estándar también son medidas de dispersión, pero se ven más afectadas por valores atípicos y por la no normalidad.

$$MAD = \text{median}(|Y_i - \text{median}(Y_i)|)$$

Para calcular la MAD podemos usar un paquete de SciPy:

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.median_abs_deviation.html

Regresión Polinómica

SKLEARN

Regresión Polinómica

Se trata de generar una matriz cuyos coeficientes sean combinaciones polinómicas de las variables.

Se logra un polinomio que ajuste mejor nuestros datos en vez de una recta.

$$y = w_0 + w_1x + w_2x^2 + \dots + w_dx^d$$

d: Grado del Polinomio

<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.PolynomialFeatures.html>

Regresiones Polinómicas

Notebook: <https://colab.research.google.com/drive/18ZhheYRTVlvcuKymd7qBjEDmw8RhJIhN?usp=sharing>

Regresiones

