

Modelos Supervisados

Regresión Lineal

Andrés Martínez

22 Junio, 2024

- Regresión Lineal.
- Gradiente Descendiente.
- Penalización.
- Métricas
- Inferencia

- La regresión lineal es un modelo estadístico que busca relacionar una variable dependiente con una o más variables independientes.
- En términos generales, la regresión es un promedio.
- En Ciencia de Datos la interpretación se concentra en la predicción.
- El gradiente descendente es un algoritmo de optimización que nos permite encontrar los coeficientes del modelo de regresión.

Regresión Lineal

Ecuación de Regresión Lineal:

$$y = \beta_1 x + \beta_0$$

- β_1 es la pendiente
- β_0 es la intersección en y-eje

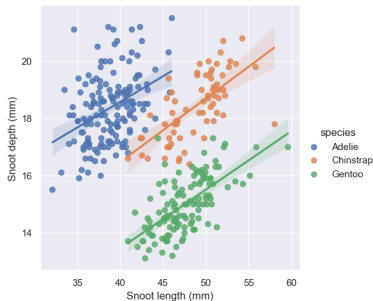


Figure 1: Regresión

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- Y es la variable dependiente.
- β_0 es el intercepto.
- β_1 es el coeficiente de regresión.
- X es la variable independiente.
- ε es el término de error.

$$J(\beta_0, \beta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\beta}(x^{(i)}) - y^{(i)})^2$$

- $J(\beta_0, \beta_1)$ es la función de costo que queremos minimizar.
- m es el número de muestras en nuestros datos.
- $h_{\beta}(x^{(i)})$ es la predicción del modelo para la muestra i .

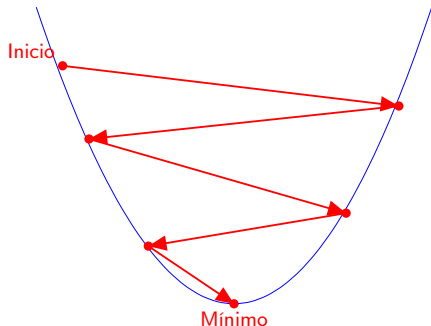
$$\beta_0 = \beta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\beta}(x^{(i)}) - y^{(i)})$$

$$\beta_1 = \beta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\beta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}$$

- α es la tasa de aprendizaje.

- Inicializamos β_0 y β_1 .
- Repetimos hasta la convergencia:
 - Calculamos las derivadas parciales de la función de costo.
 - Actualizamos β_0 y β_1 utilizando las derivadas y la tasa de aprendizaje α .

Gradiente Descendiente



Usamos el siguiente proceso iterativo para encontrar el mínimo:

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

Donde x_k es el punto actual, α es la tasa de aprendizaje y $\nabla f(x_k)$ es el gradiente de la función en x_k .

Ejemplo

Consideremos la función univariante $f(x) = 0.5(x - 3)^2 + 2$. Queremos encontrar el mínimo de esta función utilizando el método del gradiente descendente.

La derivada de $f(x)$ es $\nabla f(x) = x - 3$.

Elegimos una tasa de aprendizaje $\alpha = 0.1$ y un punto de inicio $x_0 = 6$.

$$\nabla f(x) = \frac{d}{dx} (0.5x^2 - 3x + 4)$$

$$\frac{d}{dx}(0.5x^2) = x$$

$$\frac{d}{dx}(-3x) = -3$$

$$\frac{d}{dx}(4) = 0$$

$$\nabla f(x) = x - 3 + 0 = x - 3$$

El gradiente de la función $f(x)$ es $\nabla f(x) = x - 3$.

Usamos el siguiente proceso iterativo para encontrar el mínimo:

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

Donde x_k es el punto actual, α es la tasa de aprendizaje y $\nabla f(x_k)$ es el gradiente de la función en x_k .

$$x_1 = 6 - 0.1 \cdot (6 - 3) = 6 - 0.3 = 5.7$$

$$x_2 = 5.7 - 0.1 \cdot (5.7 - 3) = 5.7 - 0.37 = 5.33$$

Continuamos con el proceso iterativo del gradiente descendente:

$$x_3 = 5.33 - 0.1 \cdot (5.33 - 3) = 5.33 - 0.233 = 5.097$$

$$x_4 = 5.097 - 0.1 \cdot (5.097 - 3) = 5.097 - 0.1797 = 4.9173$$

$$x_5 = 4.9173 - 0.1 \cdot (4.9173 - 3) = 4.9173 - 0.19173 = 4.72557$$

Después de cinco iteraciones, el método del gradiente descendente encuentra un valor cercano al mínimo de la función.

$$xf(x)^2 + 2; f(x)$$

- En la regresión lineal, queremos encontrar un modelo que se ajuste a los datos.
- Sin embargo, a menudo queremos prevenir el sobreajuste y mejorar la generalización del modelo.
- Para esto, utilizamos penalizaciones que controlan los coeficientes del modelo.

La penalización Lasso agrega un término de penalización L_1 a la función de costo de la regresión lineal:

$$\text{Costo}_{\text{Lasso}} = \text{Costo}_{\text{Regresión Lineal}} + \lambda \sum_{i=1}^n |\theta_i|$$

Donde:

- λ es el hiperparámetro de penalización.
- θ_i son los coeficientes del modelo.

Penalización Ridge y Elastic Net

La penalización Ridge agrega un término de penalización L_2 a la función de costo:

$$\text{Costo}_{\text{Ridge}} = \text{Costo}_{\text{Regresión Lineal}} + \lambda \sum_{i=1}^n \theta_i^2$$

La penalización Elastic Net combina ambas penalizaciones Lasso y Ridge:

$$\text{Costo}_{\text{Elastic Net}} = \text{Costo}_{\text{Regresión Lineal}} + \lambda_1 \sum_{i=1}^n |\theta_i| + \lambda_2 \sum_{i=1}^n \theta_i^2$$

Donde:

- λ_1 y λ_2 son hiperparámetros de penalización.

Diferencias entre Lasso y Ridge

Aspecto	Lasso	Ridge
Tipo de penalización	L_1	L_2
Función de costo	$\text{Costo}_{\text{Lasso}} = \text{Costo}_{\text{Regresión Lineal}} + \lambda \sum_{i=1}^n \theta_i $	$\text{Costo}_{\text{Ridge}} = \text{Costo}_{\text{Regresión Lineal}} + \lambda \sum_{i=1}^n \theta_i^2$
Significado de λ	Controla la cantidad de regularización	Controla la magnitud de los coeficientes
Efecto en los coeficientes	Puede llevar a coeficientes exactamente iguales a cero	Reduce la magnitud de los coeficientes, pero rara vez los vuelve cero
Selección de características	Realiza selección automática de características	No realiza selección automática de características
Uso común	Útil cuando se desea un modelo con un conjunto de características reducido	Útil para prevenir el sobreajuste y reducir la multicolinealidad

Error Cuadrático Medio (MSE)

El Error Cuadrático Medio (MSE) es una medida de la dispersión entre los valores reales y los valores predichos en un modelo. Se calcula como:

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

Donde:

n : Número de observaciones

y_i : Valor real

\hat{y}_i : Valor predicho

Error Cuadrático Medio de la Raíz (RMSE)

El Error Cuadrático Medio de la Raíz (RMSE) es simplemente la raíz cuadrada del MSE y se usa para proporcionar una medida del error en la misma unidad que la variable de respuesta. Se calcula como:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Error Porcentual Absoluto Medio (MAPE)

El Error Porcentual Absoluto Medio (MAPE) mide el error porcentual promedio en las predicciones. Se calcula como:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{y_i} \right) \times 100\%$$

Coeficiente de Determinación (R^2)

El Coeficiente de Determinación (R^2) indica cuánta variabilidad en la variable de respuesta es explicada por el modelo. Se calcula como:

$$R^2 = 1 - \frac{\text{MSE del modelo}}{\text{Varianza de los datos}}$$

Donde:

MSE del modelo : Error Cuadrático Medio del modelo

Varianza de los datos : Varianza de los valores reales (y)

- La regresión lineal es un modelo útil para predecir relaciones lineales entre variables.
- El gradiente descendente es un algoritmo eficaz para encontrar los coeficientes óptimos del modelo.