Modelos Supervisados

Certificación en Ciencia de Datos

Andrés Martínez

28 Junio 2024

Universidad Externado de Colombia

135 Años

Contenido

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

Introducción

- La regresión logística es una técnica fundamental en la ciencia de datos que se utiliza para abordar problemas de clasificación.
- Se utiliza principalmente en problemas de clasificación binaria, donde el objetivo es predecir una de las dos clases posibles, como "sí/no," "fraude/no fraude," o "enfermo/sano."
- Transforma la salida en un rango entre 0 y 1, lo que la hace adecuada para modelar probabilidades.
- Los coeficientes se traducen en log-odds (logaritmo de la razón de probabilidades), que se utilizan para predecir la probabilidad.
- Se entrena mediante el método de máxima verosimilitud. Se ajustan los coeficientes para maximizar la probabilidad de que el modelo prediga correctamente las etiquetas de clase observadas en los datos de entrenamiento.

Regresión Logística

Ecuación de Regresión Logística:

$$P(y=1) = \frac{1}{1 + e^{-(\beta_1 x + \beta_0)}}$$

- $P(y=1)$ es la probabilidad de pertenecer a la clase 1

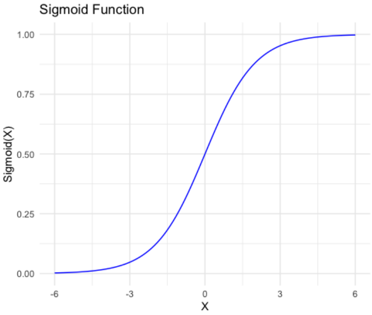

Figure 1: Regresión Logística

Regresión Logística

- Modelo de Regresión Logística

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \tag{1}$$

- Función Logística

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X \tag{2}$$

Regresión Logística (Cont.)

- Estimación de Coeficientes

$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right] \tag{3}$$

- Predicción de Probabilidades

$$\hat{p}(X) = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 X)}} \tag{4}$$

- Modelo de Regresión Logística Ordinal

$$P(Y \leq j|X) = \frac{1}{1 + e^{-(\alpha_j - \beta X)}} \tag{5}$$

- Categorías Ordenadas

$$P(Y = 1|X) < P(Y = 2|X) < \ldots < P(Y = J|X) \tag{6}$$

Regresión Logística Ordinal (Cont.)

- Estimación de Coeficientes

$$\hat{\alpha}_j, \hat{\beta} = \arg\min_{\alpha_j, \beta} \sum_{i=1}^{n} L(y_i, \hat{P}(Y \leq j|X)) \tag{7}$$

- Función de Enlace Logit

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \tag{8}$$

Regresión Logística Ordinal (Cont.)

- Predicción de Probabilidades Ordinales

$$P(Y \leq j|X) = \frac{1}{1 + e^{-(\hat{\alpha}_j - \hat{\beta} X)}} \tag{9}$$

- Clasificación Multicategórica

$$\hat{Y} = \arg\max_j P(Y \leq j|X) \tag{10}$$

Gradiente Descendente en Regresión Logística

- Objetivo: Encontrar los coeficientes β_0 y β_1 que minimizan la función de costo.
- Función de Costo (Log Loss)

$$J(\beta_0, \beta_1) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\right] \tag{11}$$

- Gradiente de la Función de Costo

$$\frac{\partial J}{\partial \beta_0} = -\frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i - y_i) \tag{12}$$

$$\frac{\partial J}{\partial \beta_1} = -\frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i - y_i)X_i \tag{13}$$

Gradiente Descendente (Cont.)

- Inicialización de Coeficientes

$$\beta_0 = 0, \quad \beta_1 = 0 \tag{14}$$

- Actualización de Coeficientes

$$\beta_0 = \beta_0 - \alpha \frac{\partial J}{\partial \beta_0} \tag{15}$$

$$\beta_1 = \beta_1 - \alpha \frac{\partial J}{\partial \beta_1} \tag{16}$$

- α es la tasa de aprendizaje (learning rate).
- Repetir las actualizaciones hasta converger o alcanzar un número máximo de iteraciones.

Tasa de Aprendizaje

- La tasa de aprendizaje (α) es un hiperparámetro crucial en el gradiente descendente.
- Un α pequeño puede hacer que la convergencia sea lenta, mientras que uno grande puede provocar divergencia.
- Elección de α debe ser cuidadosa y puede requerir ajustes y experimentación.
- La elección adecuada de α es esencial para un entrenamiento eficiente del modelo.

Iteraciones en el Gradiente Descendente

- En cada iteración, los coeficientes se actualizan en la dirección del gradiente descendente.
- El proceso se repite hasta que se alcanza la convergencia.
- Monitorear el valor de la función de costo para evaluar el progreso del algoritmo.
- El número de iteraciones necesarias puede variar según el problema y la configuración.

Métricas de Modelos de Clasificación

- En la evaluación de modelos de clasificación, las métricas calculan la tasa de exito para identificar un individuo de acuerdo a una categoría.
  - Precisión (Accuracy)
  - Recall (Sensibilidad o True Positive Rate)
  - Especificidad (True Negative Rate)
  - F1 Score (Puntuación F1)
  - Área Bajo la Curva ROC (AUC-ROC)
  - Curva ROC (Receiver Operating Characteristic)

Matriz de Confusión

La matriz de confusión es una tabla que describe el rendimiento de un modelo de clasificación.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 1: Confusion Matrix with Color Highlighting

- Verdaderos Positivos (TP): Instancias positivas clasificadas correctamente como positivas.
- Falsos Positivos (FP): Instancias negativas clasificadas incorrectamente como positivas.
- Verdaderos Negativos (TN): Instancias negativas clasificadas correctamente como negativas.
- Falsos Negativos (FN): Instancias positivas clasificadas incorrectamente como negativas.
- Estos cuatro elementos componen la matriz de confusión.

Precisión y Recall

- Precisión: Mide la proporción de verdaderos positivos entre todas las predicciones positivas.

$$\text{Precisión} = \frac{TP}{TP + FP} \tag{17}$$

- Recall: Mide la proporción de verdaderos positivos entre todas las instancias positivas en los datos reales.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

F1 Score

- El F1 Score es una métrica que combina precisión y recall en un solo número.

$$F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}} \tag{19}$$

- Proporciona un equilibrio entre precisión y recall.

AUC - Área Bajo la Curva ROC

- La curva ROC (Receiver Operating Characteristic) es una representación gráfica del rendimiento de un modelo de clasificación.
- El AUC mide el área bajo la curva ROC.
- Un valor de AUC cercano a 1 indica un buen rendimiento del modelo.

Curva ROC

- La curva ROC muestra la tasa de verdaderos positivos (recall) frente a la tasa de falsos positivos (1 - especificidad) para diferentes umbrales de clasificación.
- Un modelo perfecto tiene una curva que pasa por el punto (0,1).