

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Tecnologías de la Información: Desarrollo y Gestión de Software



Diagnostico Unidad 3: Análisis supervisado

IDGS91N - Kevin Iván Aguirre Silva

Extracción de Conocimiento en Bases de Datos - Ing.

Luis Enrique Mascote Cano

Chihuahua, Chih., 29 de noviembre de 2025

1. ¿Cuál no es un algoritmo de regresión?

A) Regresión lineal

B) K-Means

C) Lasso

D) Ridge

2. En clasificación, ¿qué métrica mide la proporción de verdaderos positivos correctamente identificados?

A) Precision

B) Recall

C) Accuracy

D) MSE

3. ¿Qué técnica ayuda a evitar el sobreajuste ajustando penalizaciones?

A) Cross-validation

B) Regularización

C) PCA

D) DBSCAN

4. ¿Cuál es el propósito de GridSearchCV?

A) Ajustar hiperparámetros

B) Dibujar curvas ROC

C) Preprocesar datos

D) Reducir dimensionalidad

5. ¿Qué representa la curva ROC?

A) Relación entre precisión y recall

B) Tasa de verdaderos positivos vs. tasa de falsos positivos

C) Error vs. tiempo

D) Correlación de variables

Respuesta Corta

6. Menciona un caso práctico donde usarías regresión.

Por ejemplo para la predicción del precio de una casa, creo que es posible basarse en características como el tamaño, número de habitaciones, etc. De esta manera se utiliza la regresión para estimar el precio de la casa a partir de todas sus variables de entrada.

7. ¿Por qué es útil la validación cruzada?

Puede ayudar a estimar que tan bien se generaliza un modelo estadístico a un conjunto de datos independientes de los cuales se utilizaron para entrenarlo.

8. Explica brevemente qué es una matriz de confusión.

Es una tabla que se suele utilizar para describir el rendimiento de un modelo de clasificación en un conjunto de datos de prueba.