Principio del formulario

Formulario de evaluación aprendizaje automático

Al implementar un modelo de clasificación para la detección de fraude se ha observado que se ha identificado correctamente 85 casos de fraude y 890 no fraudulentos. Por otro lado, el modelo marca como fraude 10 (falsos positivos) casos que no lo son y deja escapar 15 (falsos negativos) que si lo son. ¿Cuál es la exactitud de este modelo?

Tu respuesta

Exactitud= TP/(TP + FP) = 85 / (85+10) = 0,8947

Al implementar un modelo de clasificación para la detección de fraude se ha observado que se ha identificado correctamente 85 casos de fraude y 890 no fraudulentos. Por otro lado, el modelo marca como fraude 10 (falsos positivos) casos que no lo son y deja escapar 15 (falsos negativos) que si lo son. ¿Cuál es la precisión de este modelo?

Precision= ( TP + TN) / (TP +FP +TN +FN ) = (85+890) / ( 85 + 10 + 890 + 15) = 0,975

Tu respuesta

En un gran superficie se han observado que 100 clientes han comprado leche, 80 han comprado pan y 30 han comprado mantequilla de un total de 200 clientes. De los 80 clientes que han comprado pan 40 han comprado también leche y 10 han comprado leche, pan y mantequilla. ¿Cuál es la confianza de la regla {Leche, Pan} -> {Mantequilla}?

Confianza es : supp(X U Y )/supp(X)= supp ( (leche,pan) U mantequilla)/ supp(leche,pan) = 10 / 40 = 0,25

Tu respuesta

¿Qué se puede medir con el WoE (Weight of Evidence)?

**La capacidad de predecir de una agrupación o nivel de una variable**

La capacidad de predecir de un modelo

El peso de una variable utilizada en un modelo

La existencia de sobreajuste en un modelo

¿Qué se mide con el VIF?

**Multicolinealidad**

Colinealidad

La capacidad de predecir de un modelo

La velocidad de convergencia de un modelo

Se han estimado los siguientes centroides con el método de k-means: c1 = [1; 2], c2 = [-3; 0] y c3 = [4; 2]. Suponiendo que se utiliza la distancia Euclídea, ¿A que clúster pertenece el punto [-1; 2]?

**C1**

C2

C3

Ninguno

¿Cuál de las siguientes es una forma razonable de seleccionar el número de componentes (k) principales de un conjunto de datos con n dimensiones?

**Se selecciona el menor valor de k de modo que las componentes principales contengan por lo menos el 99% de la varianza de los datos**

Se selecciona k = n/2

Se utiliza el WoE

Se seleccionan los tres primeros

Al realizar un modelo de clasificación con regresión logística ¿Cuál de las siguientes afirmaciones son ciertas?

**Al añadir nuevas variables al conjunto de datos siempre se consigue un modelo mejor en el conjunto de entrenamiento y validación.**

**Al añadir muchas variables nuevas es más probable que aparezca sobreajuste (overfit) en el conjunto de entrenamiento.**

Se selecciona el menor valor de k de modo que las componentes principales contengan por lo menos el 99% de la varianza de los datos

¿Qué es un auto-codificador?

**Una red neuronal**

Un algoritmo de clustering

Un algoritmo de clasificación

Un sistema de recomendación

¿Cuáles de los siguientes métodos de aprendizaje automático son supervisados?

K-means

Gaussian mixtures

**Arboles de decisión**

**Regresión logística**

En un gran superficie se han observado que 100 clientes han comprado leche, 80 han comprado pan y 30 han comprado mantequilla de un total de 200 clientes. De los 80 clientes que han comprado pan 40 han comprado también leche y 10 han comprado leche, pan y mantequilla. ¿Cuál es el soporte de conjunto Leche y Pan?

**40/200 = 0,2**

¿En cuál de las siguientes aplicaciones se podría utilizar una algoritmo de clustering, por ejemplo k-means?

**A partir de los patrones de uso en un sito web identificar los diferentes grupos de usuarios que existen.**

A partir de las ventas de un almacén en un periodo de rebajas estimar las ventas en el siguiente.

A partir del histórico de navegación en una tienda electrónica estimar la probabilidad de que un nuevo visitante realice una compra.

**Teniendo en cuenta los datos de ventas de un almacén averiguar que productos forman grupos (por ejemplo se compran con frecuencia juntos) y por lo tanto se deben poner juntos.**