

Pregunta 1: ¿Cuáles son las principales diferencias entre los métodos envolvedores y los de filtrado?

Respuesta:

Métodos envolvedores	Métodos de filtrado
Seleccionan las características más “útiles”	Seleccionan las características más “relevantes” o “importantes”
Hacen uso de la validación cruzada	Hacen uso de estadísticas univariadas
Son computacionalmente muy intensivos	Son computacionalmente menos intensivos
Suelen proporcionar el conjunto de características de mejor comportamiento para el modelo.	Suelen proporcionar un conjunto de características más general, normalmente dan un rendimiento de predicción más bajo

Pregunta 2: Haciendo uso de Random Forest, ¿se puede predecir el precio de una acción de la bolsa de valores tomando en cuenta los datos históricos?, ¿y predecir si una persona puede vivir o morir en el hundimiento del Titanic tomando en cuenta la edad, sexo y ubicación de su cabina?

Respuesta: Con este algoritmo no se puede predecir el valor de una acción de la bolsa ya que para esto se requiere un algoritmo de regresión y este es un algoritmo de clasificación. Por otro lado, sí se podría predecir si una persona puede vivir o morir en el hundimiento del Titanic, ya que el objetivo es realizar una clasificación dadas unas características.

Pregunta 3: ¿Por qué se usan más soluciones de aprendizaje de máquina supervisados (85%) que no supervisados (15%), para la diabetes?

Respuesta: esto es porque la gran parte de las soluciones propuestas tienen como objeto la prevención y diagnóstico por medio de la predicción, para esto es bastante necesario que el aprendizaje sea supervisado, es decir, los conjuntos de datos que se usan en estas soluciones son datos de personas afectadas por la diabetes, en conclusión, se conocen las salidas.

Pregunta 4: Nombre los pasos para el descubrimiento de conocimiento a partir de una base de datos y dónde se aplica en aprendizaje de máquina

Respuesta: Primero se hace una selección de datos, estos se separan en subconjuntos, luego de esto se hace un pre procesamiento de los datos para transformarlos, una vez transformados se hace la minería de datos, dejando con esto modelos que se interpretan y se evalúan para finalmente llegar al conocimiento.

Pregunta 5:Cuál es la diferencia entre las Variables objetivo de un árbol de regresión y uno de clasificacion:

Respuesta: El árbol de regresión se utiliza para variables continuas, el árbol de clasificación se utiliza para variables categóricas.

Pregunta 6: Teniendo en cuenta que hay algoritmos o técnicas de clasificación y selección de datos, ¿Es posible predecir algún tipo de dato o característica con la técnica de minería de datos de patrones secuenciales?.

Respuesta: sí, es posible ya que la técnica de minería de datos patrones secuenciales se utiliza para extraer patrones de datos, donde los valores se entregan en una secuencia. Por lo tanto, tal técnica es adecuada para predecir por ejemplo la secuencia de medicamentos que se le recetará a un paciente.