



Universidad
Rey Juan Carlos

Práctica 2. Árboles de decisión y método KNN

Aprendizaje Automático

**Grado en Ingeniería de Robótica
Software**

Curso 2023/2024

Ejercicio 1. Árboles de decisión.

Un grupo de pescadores quiere utilizar un árbol de decisión para decidir de forma automática si salir a pescar o no en función de diversas variables: previsión, temperatura, marea y viento. La siguiente tabla presenta los ejemplos que se utilizarán para construir el árbol de decisión.

PREVISIÓN	TEMPERATURA	MAREA	VIENTO	PESCAR
NUBLADO	MEDIA	BAJA	MEDIO	SI
LLUVIOSO	ALTA	MEDIA	DEBIL	SI
NUBLADO	BAJA	MEDIA	FUERTE	NO
LLUVIOSO	MEDIA	ALTA	FUERTE	SI
SOLEADO	ALTA	ALTA	DEBIL	NO
LLUVIOSO	BAJA	BAJA	MEDIO	SI
SOLEADO	BAJA	ALTA	FUERTE	NO
SOLEADO	ALTA	ALTA	MEDIO	NO
NUBLADO	BAJA	MEDIA	FUERTE	NO
SOLEADO	MEDIA	BAJA	DEBIL	SI

Tabla 1. Datos de entrenamiento para construir el árbol de decisión.

A continuación, resuelva los siguientes apartados:

1.1. Realice un *script* (o un *Jupyter notebook*) de Python en que se genere el modelo basado en un árbol de decisión mediante funciones de alto nivel de la librería *scikit-learn*.

a) ¿Qué característica se ha seleccionado como nodo raíz? ¿Por qué cree que se ha seleccionado esa característica?

b) Indique si existe algún valor de alguna característica que permita salir a pescar independientemente del resto de valores de las demás características.

c) ¿Existe algún atributo que no influya en la decisión de salir a pescar o no?

d) Adjunte de forma gráfica el árbol de decisión que se ha generado.

1.2. En la siguiente tabla se muestran nuevos datos que se van a utilizar para evaluar el modelo creado en el apartado anterior.

PREVISIÓN	TEMPERATURA	MAREA	VIENTO	PESCAR
SOLEADO	ALTA	MEDIA	DEBIL	NO
NUBLADO	BAJA	MEDIA	MEDIO	NO
NUBLADO	MEDIA	BAJA	MEDIO	SI
SOLEADO	ALTA	ALTA	FUERTE	NO
LLUVIOSO	MEDIA	BAJA	MEDIO	SI
LLUVIOSO	MEDIA	ALTA	DEBIL	SI
NUBLADO	BAJA	MEDIA	MEDIO	SI
SOLEADO	ALTA	ALTA	DEBIL	NO
NUBLADO	MEDIA	BAJA	MEDIO	SI
LLUVIOSO	BAJA	MEDIA	FUERTE	SI

Tabla 2. Nuevos datos para evaluar el modelo de árbol de decisión generado.

1.2. Realice un nuevo *script* (o un *Jupyter notebook*) en el que se obtenga el porcentaje de acierto de los nuevos datos aplicando el modelo que se ha generado. De acuerdo con el dato de porcentaje obtenido, ¿Cree que el modelo será útil para el grupo de pescadores?

1.3. Genere en un nuevo *script* (o un *Jupyter notebook*) el modelo basado en árbol de decisión utilizando los datos del ejercicio 3 de la práctica 1 (regresión logística). Adjunte en la memoria una imagen del modelo. De acuerdo con el modelo generado, ¿cuál es el atributo que presenta una menor entropía?

1.4. A continuación calcule el error en porcentaje que presenta el modelo sobre los datos de entrenamiento. ¿Por qué cree que existe esa diferencia con respecto a utilizar un modelo basado en regresión logística?

Ejercicio 2. Método KNN.

Se dispone de un conjunto de datos de entrenamiento compuesto por 1000 ejemplos que se almacenan en los archivos *data.csv*.

Resuelva los siguientes apartados:

2.1. Realice un *script* (o un *Jupyter notebook*) en el que se carguen los datos de entrenamiento y se representen en una figura. Atendiendo a lo que observa, ¿Cree que el método KNN es el más apropiado para generar el modelo de clasificación? En caso afirmativo justifique su respuesta y en caso negativo, proponga otro método de los que se han estudiado en clase y justifique qué ventaja presentaría con respecto al método KNN.

2.2. Genere el modelo KNN utilizando las funciones de alto nivel de la librería *scikit-learn* con el conjunto de ejemplos de entrenamiento proporcionados y con los argumentos que considere oportunos para clasificar el vector $x = (2, 1, 3)$. ¿A qué clase pertenece si se tiene en cuenta cómo están distribuidas las clases? Argumente su respuesta.

2.3. Calcule la matriz de covarianza de cada una de las clases utilizando el conjunto de datos de entrenamiento. Atendiendo a los valores de las matrices de covarianza, ¿Observa alguna diferencia significativa entre clases? Razone su respuesta.