



UNIVERSITAS
Miguel Hernández

Tema 6. Random Forest. Ejemplos.

José L. Sainz-Pardo Auñón

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones.

1. Carga y Preparación de los Datos

- Descargar el archivo `clasificacion.xlsx` que contiene los datos sobre los que deseamos emplear la técnica de clasificación.
- Leer el archivo Excel en un DataFrame de pandas.
- Visualizar las primeras filas del conjunto de datos.
- Definir las variables independientes X_1 , X_2 , X_3 y la variable dependiente Y .

2. Obtención del modelo.

- Divide los datos en un conjunto test del 30% y uno de entrenamiento del 70%.
- Utiliza Random Forest (`RandomForestClassifier`) para clasificar los individuos de dicha base de datos.
- Configura parámetros como `n_estimators` (número de árboles) y `max_features` (número de variables aleatorias).

3. Evaluación del modelo

- Obtén los pronósticos de la muestra de prueba según el modelo de Random Forest.
- Obtén la tabla de confusión.
- Obtén el informe de clasificación del modelo utilizando la librería `sklearn`.

4. Interpretación.

- Dado que un Random Forest es una combinación de varios árboles, interpreta la importancia de las variables.
- Utiliza `feature_importances_` para identificar las variables más relevantes para la clasificación.

5. Parámetros.

En la función `RandomForestClassifier` pueden configurarse distintos parámetros para mejorar el modelo:

- `n_estimators`: el número de árboles en el bosque.
- `max_features`: el número de variables seleccionadas en cada división.
- `min_samples_split`: el número mínimo de muestras necesarias para dividir un nodo.
- `min_samples_leaf`: el número mínimo de muestras necesarias en una hoja.

Experimenta con distintas configuraciones hasta obtener el modelo que consideres más relevante.

6. Random Forest aplicado a regresión.

- Realiza una regresión sobre el fichero `regresion.xlsx` utilizando `RandomForestRegressor`.
- Obtén el error cuadrático medio (MSE) y el R^2 del modelo.
- Genera un gráfico de los valores observados frente a los valores predichos.



UNIVERSITAS
Miguel Hernández