

UNIDAD TEMÁTICA 4: Selección de Atributos

Trabajo de Aplicación 10

Escenario – Ejercicio 0

Utilizaremos para todos los ejercicios el mismo dataset, Sonar (Iris). El dataset tiene 208 ejemplos, donde cada uno representa el espectro de frecuencia del retorno de objetos en el agua. Se tiene en total 60 bandas de frecuencia diferentes, que componen los atributos del dataset.

- Carga el dataset y realiza una visualización como plot paralelo.
- Cada línea en el gráfico representa un ejemplo del dataset. El color (rojo, azul) indica si el objeto correspondiente es una roca o una mina.
- Puedes observar que hay algunas áreas del espectro de frecuencias en donde las dos clases están más diferenciadas.
- Cambia el tipo de gráfico a “deviation”. En éste sólo se muestran los valores medios para cada atributo, en ambas clases. Las secciones transparentes representan los desvíos estándar de los atributos para cada clase.
- Se puede aquí apreciar cuatro áreas del espectro donde las clases son más diferentes.
- Lista los atributos contenidos en cada una de las 4 áreas.

Utilizaremos Naive Bayes como modelo de “benchmarking”, y aplicaremos el mismo cross-validation con 5 plegamientos en todos los casos (para esto precisamos especificar la misma “semilla” en todos los modelos)

Ejercicio 1 - línea de base (benchmarking)

- Crea un nuevo modelo, “línea base”
- Agrega un “retrieve” con el dataset Sonar
- Agrega un subproceso “cross validation”, con 5 plegamientos, muestreo estratificado y una semilla = 1992.
- Dentro del subproceso agrega del lado izquierdo un bloque Naive Bayes (con corrección lapaciana), y del lado derecho los bloques “Apply Model” y “Performance” (clasification)
- Ejecuta el proceso y documenta los resultados obtenidos.