

Tema 8 (II)

Árboles de decisión y métodos de ensamble

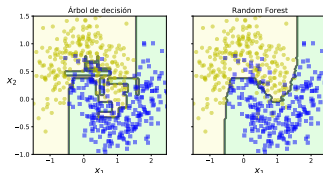
Técnicas multivariantes

Dr. Antoni Ferragut



- Agrupan predictores para obtener mejores predicciones de las que se obtendrían por separado
- Dos estrategias:
 - Escoger métodos muy diferentes entre sí
 - Usar el mismo método con diferentes muestras de entrenamiento: **bagging** (con reemplazamiento), **pasting** (sin reemplazamiento)
- El bagging permite que una misma observación se utilice varias veces para el mismo modelo predictor
- El ensamble permite realizar predicciones de una nueva observación agregando (moda/media) las predicciones de cada uno de los predictores que lo conforman
- *scikit-learn BaggingClassifier (bootstrap = True/False)*

- Utiliza bagging (o menos frecuentemente pasting) y cada uno de los predictores es un árbol de decisión
- Se generan fijando el número máximo de nodos hoja a un valor razonablemente pequeño
- Se escoge la mejor variable de una selección aleatoria de todas las variables disponibles
- *scikit-learn RandomForestClassifier*
- Es posible evaluar la importancia de las variables predictoras en el modelo de ensamble cuantificando cómo contribuyen a reducir la impureza en promedio entre los diferentes árboles de decisión



- Consisten en el ensamble de predictores débiles para lograr un predictor fuerte
- Se trata de entrenar estos predictores de forma secuencial, tratando de que cada nuevo predictor vaya mejorando a su predecesor
- Boosting adaptativo (Adaboost), Boosting por gradiente (Gradient Boosting)

- Presta más atención a las observaciones de entrenamiento que el predecesor ha ajustado erróneamente
- Los sucesivos predictores van ajustando las observaciones más difíciles de clasificar, ajustando pesos a las observaciones
- Permite realizar una predicción agregando los resultados, ponderando los predictores según los pesos en la clasificación

- Ajusta los predictores de forma secuencial corrigiendo cada nuevo predictor a su predecesor
- Ajusta cada nuevo predictor con los residuos del predictor previo
- Es necesario seleccionar:
 - La profundidad máxima permitida, d , de cada uno de los árboles de decisión
 - El número máximo de árboles, N , es decir, de predictores en el ensamble
 - La tasa de aprendizaje que va a determinar la velocidad a la cual se van a ir incorporando los nuevos predictores a los residuos