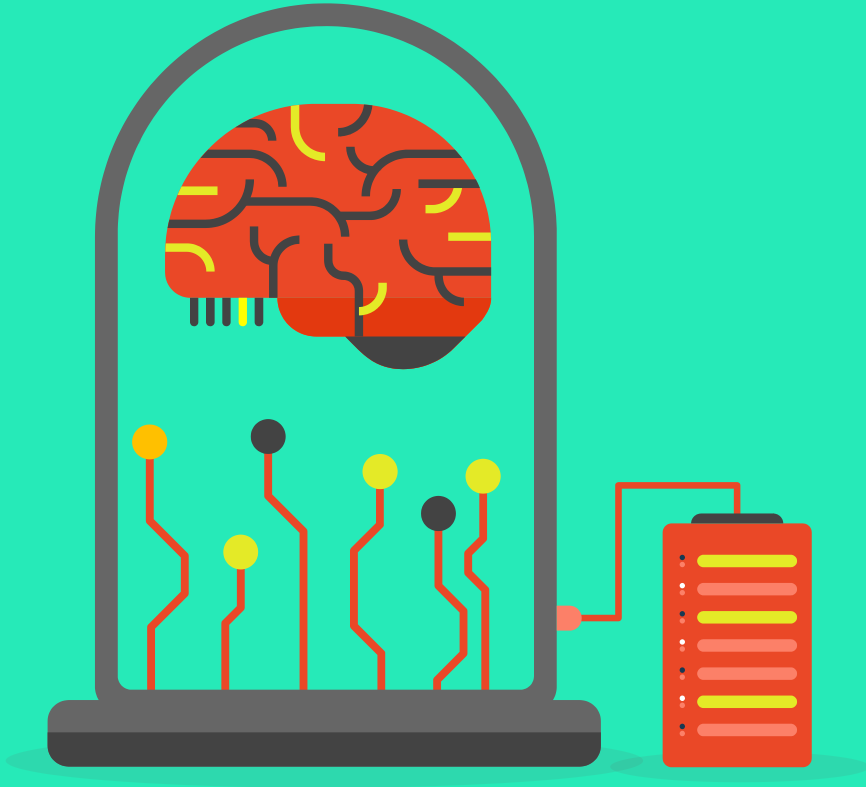


13. Clasificación con Árboles de Decisión



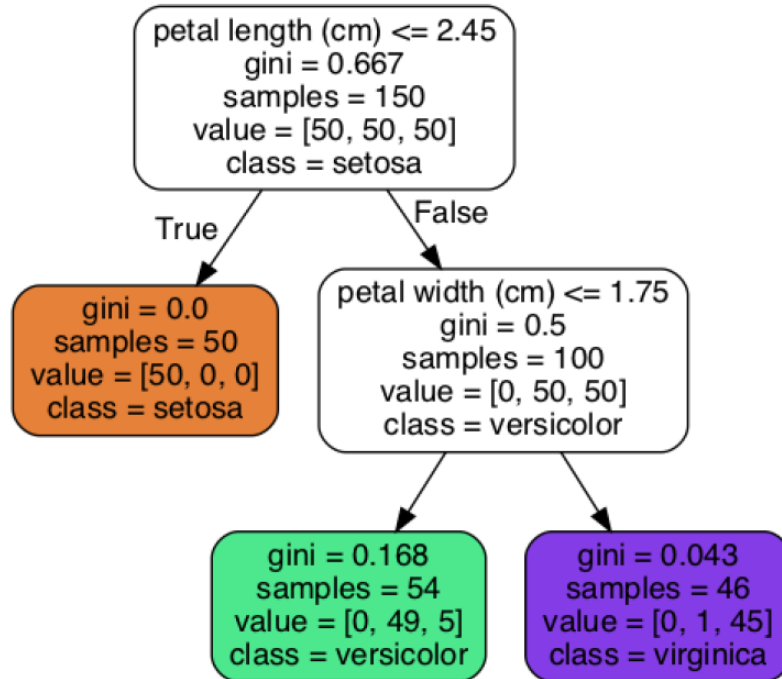
Árboles de Decisión

Como ya mencionamos anteriormente, los **Árboles de Decisión** son uno de los algoritmos de Machine Learning más versátiles que existen, aplicables tanto a problemas de Regresión como de Clasificación, y que representan el punto de partida para los modelos de Bosques Aleatorios.

Recordemos, entonces, la forma que tenía el árbol de decisión de ejemplo capaz de resolver el problema de clasificación del dataset de Iris.

Árboles de Decisión

En este caso, entendemos ahora claramente el significado de las clases al final de cada rama del árbol:



Árboles de Decisión

Para la clasificación, el entrenamiento parte de los mismos principios explicados en la sección pasada, en donde se aplica el algoritmo CART y se minimiza la función de costo basada en impurezas de las divisiones del árbol en cada nivel de profundidad del mismo. Las técnicas de regularización se mantienen, y la diferencia fundamental con el caso de regresión es que el valor predicho será ahora una clase y no un valor numérico dado.

Las métricas para la evaluación del desempeño (matriz de confusión, exactitud, precisión, etc) se aplican del mismo modo para este tipo de algoritmos, y se aplican los mismos hiperparámetros a fin de evitar el sobre entrenamiento.

De manera que ya podemos implementar un ejemplo para clasificación con los conocimientos que tenemos sobre este tipo de modelos de Machine Learning.

Décimo Cuarto Notebook Práctico

Implementemos la clasificación a partir de árboles de decisión, aplicado sobre el mismo conjunto de datos que trabajamos en la práctica pasada, el UCI Wine Dataset.

Décimo Cuarto Notebook Práctico