

Control 2

1) [1 pts] *White box* vs. *Black box*

Vamos a ver más adelante que los modelos *Random Forest* y de Redes Neuronales se tienen a considerar de caja negra (*black box*) porque si bien hacen muy buenas predicciones, a veces no es fácil explicar sus decisiones. Por ejemplo, si uno de estos modelos reconoce que en una imagen hay un perro, no es fácil saber si reconoció sus orejas, su forma, su cola, etc.

Por otro lado, un árbol de decisión se tiende a considerar un modelo *white box*, porque sus decisiones son fáciles de explicar.

- Comenta por qué crees que las decisiones que toma un árbol de decisión son consideradas fáciles de explicar.

2) [2 pts] Árboles de regresión y *overfitting*

En clases vimos dos parámetros que uno le puede entregar a un árbol de decisión en **Scikit Learn**; estos eran `max_depth` y `min_samples_leaf`. El primero nos dice la profundidad máxima que puede tener el árbol. El segundo nos dice el número mínimo de elementos que debe tener una hoja al momento de hacer *split*. En el contexto de árboles de regresión, responde:

- ¿Por qué limitar la profundidad puede prevenir el *overfitting*?
- ¿Por qué entregar un mínimo al número de elementos para hacer *split* puede prevenir el *overfitting*?

Se espera que analice cada caso por separado, dando un ejemplo convincente.

3) [3 pts] El truco del kernel

En el contexto de SVM discutimos una técnica llamada “el truco del Kernel” (*The Kernel Trick*). En ese contexto se pide que expliques:

- ¿Para qué sirve el truco del Kernel?
- ¿Por qué este truco se puede usar junto al modelo de SVM?
- Explique en qué consiste el kernel polinomial y el kernel *gaussian rbf*.