

## Examen ML – 30/06/22

Método: contestar las siguientes preguntas en su propia escritura, en un plazo no mayor a 3 horas, a “libro cerrado”. Al terminar escanear las páginas escritas y mandarlas por mail.

1. Espacio de hipótesis y búsqueda: Explique brevemente estos dos conceptos. Describa el espacio de hipótesis y el tipo de búsqueda del algoritmo ID3 (árboles de decisión) en el problema de las espirales anidadas.
2. Nombre al menos dos criterios con los cuales se detiene el crecimiento de un árbol de decisión durante el entrenamiento, o ajuste, del mismo.
3. Al entrenar redes neuronales dividimos típicamente nuestros datos en 3 conjuntos: entrenamiento, validación y test. Explique para qué se usa cada uno.
4. Describa dos diferencias importantes entre el Aprendizaje Bayesiano y los demás métodos que analizamos en el curso.
5. Discuta el problema que produce usar el error de clasificación “normal” (la proporción de puntos incorrectamente asignados) para evaluar un clasificador en un problema con dos clases muy desbalanceadas.
6. Vimos en el curso que usando un número mayor de vecinos podemos lograr que el clasificador de k-vecinos funcione bien en problemas ruidosos, como son los datos “diagonal” y “paralelo”. Cuál es el problema que impide que pase lo mismo con el dataset de “espirales con ruido”?
7. Cómo hacen las SVM para encontrar soluciones a problemas de clasificación con ruido, donde algunos pocos puntos no se pueden separar correctamente con un hiperplano? Y si directamente no se puede solucionar el problema con un hiperplano, como el XOR?