

OPENCOURSEWARE
APRENDIZAJE AUTOMÁTICO PARA EL ANÁLISIS DE DATOS
GRADO EN ESTADÍSTICA Y EMPRESA
Ricardo Aler



1. El método de Sequential Model-based Optimization (SMO) construye un modelo de regresión a partir de puntos explorados previamente. ¿Para qué usa SMO ese modelo y cómo?
2. Si f es una neurona de una red de neuronas y tiene 3 entradas x_1 , x_2 , x_3 , ¿cuales serían las operaciones matemáticas f que realiza la neurona para computar la salida y ? ($y=f(x_1, x_2, x_3)$).
3. ¿Boosting puede tener problemas de sobreaprendizaje al ruido en problemas de clasificación? Explicar la respuesta.
4. Supongamos que tenemos un conjunto de datos disponibles D y queremos construir y evaluar un modelo usando sólo los atributos más relevantes. Supongamos que seguimos el siguiente proceso: 1) usamos un método de selección de atributos usando el conjunto D y elegimos los mejores; 2) después dividimos D en una partición de entrenamiento y otra de test; 3) construimos el modelo con la de entrenamiento y lo evaluamos con la de test. ¿Es correcto este proceso? ¿Por qué?
5. ¿En qué se parecen y en qué se diferencian los Extremely-randomized trees de los Random Forests?

6. Supongamos que tenemos unos datos disponibles D para un problema de clasificación bi-clase, y queremos construir dos particiones ESTRATIFICADAS: entrenamiento ($\frac{2}{3}$) y test ($\frac{1}{3}$). Describir el procedimiento que habría que utilizar para construir esas particiones.

7. El proceso de entrenamiento de una SVM lineal involucra minimizar esta expresión: $(1/\text{margen}^2) + C \cdot \sum \xi_i$. ¿Qué le pasará a una SVM si hacemos que el hiper-parámetro C sea muy grande? ¿Por qué?

8. Supongamos que queremos utilizar Map/Reduce para contar cuantas veces aparece cada palabra en un fichero que contiene todos los textos de la literatura universal. Explicar con palabras el proceso que se seguiría para contar palabras y en concreto, qué es lo que tendría que hacer la función Map, qué la función Reduce y qué sort-and-shuffle.

9. En el algoritmo del vecino más cercano, para qué valores de K (número de vecinos) se puede producir más sobre-aprendizaje? ¿para valores pequeños o para valores grandes? ¿Por qué?

10. Supongamos que un clasificador trivial clasifica las instancias como positivas con probabilidad 0.9 y como negativas con probabilidad 0.1. ¿Cuál sería el True Negative Rate de este clasificador? ¿Por qué?