

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



EXAMEN DE APRENDIZAJE AUTOMÁTICO
GRADO EN ESTADÍSTICA Y EMPRESA
10 puntos, 1 hora de duración.

Responder cada pregunta con respuestas breves (unas pocas líneas).

1. ¿Cuál es la diferencia entre clasificación y clustering?

Respuesta: la clasificación es un problema supervisado. Cada dato de entrenamiento está etiquetado con su clase (de ahí la “supervisión”) o atributo de salida y el objetivo es encontrar un modelo que dados los atributos de entrada devuelva la clase correcta. Por ejemplo, clasificar a un cliente en deudor o no deudor. El clustering por el contrario es un tipo de problemas no supervisados, los datos no están etiquetados y el objetivo es encontrar agrupaciones naturales de los datos en términos de los atributos de entrada. Varios datos pertenecerán al mismo cluster si se parecen entre ellos y se diferencian de los datos de otros clusters.

2. ¿Cuál es la diferencia entre un clasificador discreto, un scorer y un estimador de probabilidades?

Respuesta: un clasificador discreto devuelve una de entre un número finito y pequeño de clases (ejemplo: deudor / no deudor). Un scorer evalúa el grado de pertenencia (score) del dato a la clase, por ejemplo devolviendo la distancia de separación del dato a la frontera de separación entre las clases. Pero ese score no tiene porque seguir los axiomas de las probabilidades. Si el score es la probabilidad condicional de que el dato pertenezca a la clase, entonces se habla de estimación de probabilidades.

3. ¿Cuál es la diferencia entre “one versus all” y “one versus one” en el aprendizaje con mas de dos clases?

Respuesta: ambos esquemas sirven para atacar problemas multiclase mediante clasificadores biclase. One versus all construye N clasificadores biclase para separar la clase i de todas las demás. Si un dato es clasificado por distintos clasificadores como perteneciente a su clase, se usa un score para desempatar. En one versus one se construyen tantos clasificadores como parejas de clases. Gana la clase que es votada mas veces. Como se puede comprobar, one versus one requiere de la construcción de muchos mas clasificadores y no suele ser práctico para problemas con muchas clases.

4. ¿En qué consiste la validación cruzada?

Respuesta: La validación cruzada consiste en dividir el conjunto de datos en k partes y repetir un ciclo en el que se entrena con k-1 partes y se hace el test con la parte restante. Al final, se calcula la media de los k tests. Es preferible al método de entrenamiento/test porque con un sólo conjunto de test pueden aparecer sesgos por azar, mientras que con k conjuntos de test, esto es más difícil que ocurra.

5. ¿Qué es la “maldición de la dimensionalidad”?

Respuesta: a medida que crece el número de atributos de entrada de un problema (su dimensión), el número de datos que son necesarios para aproximar las fronteras de separación entre clases crece de manera no lineal (por ejemplo, en el caso de una hiperesfera, la superficie crece de manera exponencial). En términos prácticos, puede ocurrir que incluso aunque todos los atributos sean relevantes, un número escaso de datos en relación al número de atributos, puede hacer que no se aproxime correctamente la frontera de separación entre clases.

6. ¿Cuál es la principal desventaja de PCA si lo usamos para seleccionar atributos en clasificación?

Respuesta: PCA es un método no supervisado (no tiene en cuenta la clase). Aunque es útil para reducir la dimensionalidad y conseguir atributos no correlacionados, puede ocurrir que los atributos seleccionados no sean los mejores para clasificar.

7. Describir brevemente lo que hacen los dos tipos de técnicas (edición y condensación) que existen para seleccionar instancias para el algoritmo de k-vecinos

Respuesta: los métodos de edición y condensación intentan mejorar la eficiencia y calidad de los modelos KNN. La edición intenta deshacerse de aquellos datos ruidosos y dejar las fronteras de separación entre clases más limpias. La condensación intenta eliminar aquellos datos irrelevantes para la clasificación (por ejemplo, aquellos que están en el interior de la masa de datos, puesto que los que interesan son los de la frontera)

8. Describir brevemente cómo funciona LVQ

Respuesta: muy brevemente, para cada dato de entrenamiento, LVQ compara el prototipo más cercano al dato y lo aproxima cierta distancia si la clase del dato y el prototipo coinciden, y lo aleja si no coinciden.

9. ¿Qué hacen los clasificadores triviales de una curva ROC, en qué coordenadas de la curva ROC están situados y porqué?

Respuesta: los clasificadores triviales siempre responden “clase negativa” o “clase positiva”, independientemente de los atributos de entrada del dato. Sus coordenadas en la ROC son: $FP=0, TP=0$ y $FP=1, TP=1$. La razón es que si se clasifican todos los datos como negativos, nunca habrá true positives (TP), pero tampoco false positives (FP). Igualmente, si se clasifican todos como positivos, todos los positivos se clasificarán correctamente ($TP=1$), pero a costa de clasificar todos los negativos de manera errónea ($FP=1$).

10. Supongamos que tenemos tres clasificadores distintos a, b, y c. ¿Hay alguno de ellos que sea irrelevante? ¿Porqué?
- Un árbol de decisión tiene True Positives = 0.9, False Positives = 0
 - Un KNN tiene $TP=1$ y $FP=0.25$
 - Una red de neuronas tiene $TP=0.8$ y $FP=0.2$

Respuesta: los clasificadores irrelevantes se caracterizan por estar en la diagonal de la ROC, es decir $TP=FP$. Es decir, el clasificador acierta más en los positivos, pero a costa de fallar la misma proporción en los negativos. Dado que ni a, ni b, ni c cumplen con $TP=FP$, ninguno de ellos es trivial.