

Métodos Probabilísticos en Inteligencia Artificial

Septiembre 2000. Examen Teórico. 8 puntos

1. El algoritmo general de inducción de Árboles de Clasificación –TDIDT *Top Down Induction of Decision Trees*– es, desde un punto de vista de optimización, un optimizador local, ya que en cada paso busca la mejor variable con la que particionar el conjunto de datos pertenecientes a una determinada rama del árbol.

Se pide diseñar un Algoritmo Genético (función de coste, representación de individuos, operadores de cruce y mutación) con el que se supere la característica de optimalidad local del TDIDT.

2. Enuncia un algoritmo de Clasificación Ascendente Jerárquica. Aplícalo a un ejemplo de tu invención con cinco objetos tridimensionales.
3. Supongamos que en una determinada carrera universitaria el 30 por ciento de los individuos que la comienzan son capaces de terminarla en 6 o menos años.

Además se conoce que el 75 por ciento de los individuos que la terminan en 6 o menos años tuvieron una nota de selectividad superior o igual a 8 puntos. Por otra parte tan sólo el 20 por ciento de los individuos que necesitaron más de 6 años para terminarla entraron en la Universidad con una nota de selectividad superior o igual a 8 puntos.

Si tenemos en cuenta el número de créditos obtenidos durante el primer año de estancia en la Universidad, obtenemos que el 85 por ciento de los individuos que necesitaron 6 o menos años para terminar la carrera, fueron capaces de conseguir al menos 30 créditos, mientras que tan sólo el 25 por ciento de los individuos que necesitaron más de 6 años para finalizar sus estudios universitarios, fueron capaces de conseguir al menos 30 créditos.

Suponemos que los resultados proporcionados por las dos pruebas anteriores son, una vez conocido el status del individuo (≤ 6 años, > 6 años) independientes.

Conocidos los datos de dos individuos, Juan y Maria, al finalizar su primer año de estancia universitaria:

- Juan: nota de selectividad 6.8, 25 créditos en su primer año
- Maria: nota de selectividad 8.1, 35 créditos en su primer año

se quiere calcular, para cada uno de ellos, la probabilidad de terminar la carrera en 6 o menos años.

Explicar las características teóricas fundamentales del paradigma clasificatorio utilizado.

4. La tabla adjunta contiene 8 casos bidimensionales que constituyen el conjunto de entrenamiento para un clasificador k -NN. Los cuatro últimos casos -numerados del 9 al 12- forman el conjunto de testeo. Construye el clasificador k -NN a partir de la distancia euclídea y de dos valores distintos de k ($k = 1, 2$).
5. La tabla adjunta muestra los resultados obtenidos en un examen, Y (*aprobar* versus *suspender*), por una muestra de 100 alumnos, teniendo en cuenta el número de horas estudiadas, X (< 30 horas, versus ≥ 30 horas).

A un nivel de significación de $\alpha = 0.05$, testar la hipótesis de independencia entre las variables X e Y .

<i>Caso</i>	<i>X</i>	<i>Y</i>	<i>Clase</i>
1	1	1	0
2	2	4	0
3	3	2	0
4	3	5	0
5	4	4	0
6	4	7	1
7	6	4	1
8	6	6	1
9	2	7	?
10	3	4	?
11	4	6	?
12	5	5	?

	<i>< 30 horas</i>	<i>≥ 30 horas</i>
<i>aprobar</i>	10	60
<i>suspender</i>	20	10

6. Se extrae una muestra aleatoria de tamaño n de una distribución normal, X , con parámetros μ y σ , ambos desconocidos. Obtener los estimadores máximo verosímiles para dichos parámetros.

Nota. Los ejercicios 1, 2, 3 y 4 se puntúan sobre 1.5 puntos. Los ejercicios 5 y 6 sobre 1 punto. Los 2 puntos restantes provienen de la calificación de la práctica.