

Métodos Probabilísticos en Inteligencia Artificial

Septiembre 2001. Examen Teórico. 6 puntos

1. Se extrae una muestra aleatoria de tamaño n de una distribución binomial, X , con parámetros m y p , siendo este último desconocido. Obtener el estimador máximo verosímil para p . Demostrar que dicho estimador máximo verosímil es también eficiente.
2. Los 300 ordenadores personales comprados por una institución son de tres marcas distintas que denotamos por A , B y C . De cada una de las marcas se han comprado 100 aparatos. En relación a las averías graves que presentan dichos aparatos en los dos primeros años de utilización obtenemos que 20 aparatos de la marca A , 30 de la marca B y 40 de la marca C han presentado ese tipo de avería. A un nivel de significación de $\alpha = 0.05$ nos preguntamos si podemos seguir manteniendo la hipótesis de independencia entre las variables marca de aparato y avería grave en los dos primeros años.
3. La tabla adjunta contiene 6 casos bidimensionales que constituyen el conjunto de entrenamiento para un clasificador k -NN. Los dos últimos casos -numerados como 7 y 8- forman el conjunto de testeo.

Caso	X	Y	Clase
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	?
8	6	6	?

Construye el clasificador k -NN a partir de la distancia euclídea y de dos valores distintos de k ($k = 1, 2$).

4. Diseñar un algoritmo genético y un algoritmo de estimación de distribuciones para resolver el problema de la mochila. Explicar las diferencias y similitudes entre ambos, así como la adecuación de cada uno de los algoritmos al problema planteado.
5. Se conoce que el 40 por ciento de los alumnos aprueban la asignatura A en la primera convocatoria.

Supongamos que en el modelo que hemos construido para predecir si un alumno va a aprobar la asignatura A en su primera convocatoria tenemos dos variables predictivas que se suponen independientes dado el valor de la clase. Las dos variables predictivas son *aprobar en primera convocatoria la asignatura B* y *aprobar en primera convocatoria la asignatura C*. Ambas asignaturas B y C deben de ser cursadas en un curso anterior a la asignatura A siendo además requisito de la misma.

Entre los que aprueban en primera convocatoria la asignatura A el 75 por ciento había aprobado también en primera convocatoria la asignatura B , mientras que entre los que suspenden en primera convocatoria la asignatura A tan sólo el 10 por ciento había aprobado en primera convocatoria la asignatura B .

Por lo que respecta a la asignatura C , se tiene que entre los que aprueban en primera convocatoria la asignatura A el 55 por ciento habia aprobado tambien en primera convocatoria la asignatura C , mientras que entre los que suspenden en primera convocatoria la asignatura A tan sólo el 20 por ciento habia aprobado en primera convocatoria la asignatura C .

Efectuar el diagnóstico para un alumno que aprobó la asignatura B en primera convocatoria, pero necesitó mas de una convocatoria para superar la asignatura C .

6. Explicar las diferencias y similitudes entre la clasificación no supervisada particional y la clasificación ascendente jerárquica tanto por medio de un ejemplo como escribiendo el pseudocódigo de algoritmos que efectúen ambos tipos de clasificación no supervisada.