

Métodos Matemáticos en Ciencias de la Computación

1 Junio 2004. Parte Teórica. 5 puntos

1. Un problema importante en el momento de representar un grafo dado consiste en colocar los nodos del mismo de tal manera que se minimice el número de cortes entre las aristas. Vamos a considerar una versión reducida de dicho problema consistente en que los n nodos de los que consta el grafo se encuentran situados en una circunferencia, cada uno de ellos a $\frac{360}{n}$ grados del anterior.

Se pide diseñar un algoritmo genético –especificando la representación de los individuos, la función de coste, al igual que los operadores de cruce y mutación– para resolver el problema anterior.

2. Sean X e Y dos variables aleatorias discretas. Denotamos por $H(X)$ a la entropía de la variable X . De manera análoga se denota por $H(X|Y)$ a la entropía de la variable X dada la variable Y . Por medio de $D_{KL}(p(x, y), p(x) \cdot p(y))$ denotamos la divergencia de Kullback–Leibler entre la distribución de probabilidad de la variable aleatoria bidimensional (X, Y) y el producto de las distribuciones de probabilidad marginales de las variables X e Y .

Demostrar que se verifica:

$$H(X) - H(X|Y) = D_{KL}(p(x, y), p(x) \cdot p(y))$$

3. A partir de datos de otros cursos académicos conocemos que tan sólo el 40 por ciento de los individuos que matriculan su proyecto fin de carrera (pfc) son capaces de defenderlo en Junio de dicho curso. Además se sabe que de entre los que son capaces de defender su pfc en Junio el 70 por ciento se matriculan de menos de 25 créditos. Por otra parte de entre los que no son capaces de defender su pfc en Junio tan sólo el 10 por ciento tenían una carga inferior a 25 créditos. Aparte de esta información se conoce también que el número de horas dedicadas al proyecto durante el primer cuatrimestre por un individuo que si que defiende su pfc en Junio sigue un modelo de densidad normal con media 100 y desviación típica 10, mientras que para los individuos que no son capaces de defenderlo en Junio el modelo es también normal de media 50 y desviación típica 20.

Se trata de predecir si un individuo con 30 créditos matriculados y con una dedicación al proyecto durante el primer cuatrimestre de 90 horas va a ser capaz de presentar el pfc en Junio.

Explica las características del paradigma de clasificación usado indicando si has necesitado efectuar algún tipo de premisa previa.

4. A los datos de la tabla adjunta, aplícale:
 - (a) una edición de Wilson
 - (b) una condensación de Hart

<i>Caso</i>	X_1	X_2	<i>Clase</i>
1	1	1	0
2	2	4	1
3	3	2	0
4	3	5	0
5	4	4	0
6	4	7	1
7	6	4	1
8	6	6	1

5. Con un ejemplo de tu invención muestra que los resultados obtenidos por los algoritmos de clasificación no supervisada, Forgy y McQueen, son distintos. Explica las analogías y las diferencias entre ambos algoritmos.