

Métodos Matemáticos en Ciencias de la Computación

05 Junio 2002. Parte Teórica. 6 puntos

1. Dadas las limitaciones de los dos algoritmos de clasificación no supervisada vistos en clase (Forgy y McQueen) acerca de la optimización local por la que se rige el proceso de búsqueda de la mejor partición de un conjunto de objetos, diseñar un algoritmo genético para tal problema. Especificar la representación de los individuos, la función de coste, al igual que los operadores de cruce y mutación.
2. Por una carretera pasan vehículos siguiendo un proceso de Poisson con parámetro $\lambda = 40$ vehículos por minuto. Teniendo esto en cuenta responde a los siguientes apartados:
 - Dibuja una figura que muestre las distintas familias de variables que aparecen en un proceso de conteo y explica las propiedades y características de cada una de ellas.
 - Expresa la fórmula, sin necesidad de realizar los cálculos, de la probabilidad de que hasta el minuto 5 hayan pasado como mucho 400 vehículos.
 - ¿Cuál es la probabilidad de que el tiempo entre el vehículo séptimo y el octavo se exceda en 3 minutos? ¿Cuándo esperamos que pase el vehículo 837?
 - Suponed que fuera de la carretera hay una persona viendo pasar los vehículos, la cual enciende un cigarro. Asumiendo que el tiempo que tarda en fumarse dicho cigarro está exponencialmente distribuido con media 4 minutos, y dado un momento aleatorio, ¿cuál es la probabilidad de que pase un vehículo antes de que acabe el cigarro?
 - Si cada vehículo que pasa es con probabilidad $\frac{1}{4}$ una furgoneta y con probabilidad $\frac{3}{4}$ un turismo, entonces:
 - Calcula la probabilidad de que pasen 80 furgonetas entre los minutos 5 y 10.
 - Si entre el paso de la quinta y sexta furgonetas han transcurrido 3 minutos, calcula la probabilidad con la que el tiempo entre la sexta y la séptima furgoneta sea mayor que 2 minutos.
3. Una variable aleatoria X sigue una distribución exponencial de parámetro θ , si su función de densidad es de la forma siguiente: $f_X(x; \theta) = \theta \cdot e^{-\theta t}$ con $t \geq 0, \theta > 0$. A partir de una muestra aleatoria de tamaño n extraída de X , obtener el estimador máximo verosímil para el parámetro θ .
4. Los 360 alumnos matriculados por vez primera durante este curso en la asignatura A han obtenido los siguientes resultados en su convocatoria de junio:
 - 175 no presentados
 - 95 suspensos
 - 61 aprobados
 - 19 notables
 - 10 sobresalientes

A un nivel de significación de $\alpha = 0.05$ nos preguntamos si podemos seguir manteniendo la hipótesis de que la distribución de probabilidad de la que se han obtenido los datos es una polinomial con parámetros: $p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = \frac{1}{6}, p_4 = \frac{1}{18}, p_5 = \frac{1}{36}$.

5. Con un ejemplo de tu invención en el que haya dos variables predictoras –una de ellas discreta y la otra continua– explica la utilización del paradigma de clasificación supervisada denominado naive–Bayes.
6. La tabla adjunta contiene 8 casos bidimensionales que constituyen el conjunto de entrenamiento para un clasificador k -NN.

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

- Aplica la edición de Wilson para reducir el conjunto de entrenamiento.
- Usando el método de condensación de Hart, reduce el conjunto de entrenamiento.
- Clasifica el caso (2.5, 2.5) a partir de los conjuntos de entrenamiento reducidos obtenidos anteriormente y considerando $k = 1$.