

Métodos Matemáticos en Ciencias de la Computación

17 Septiembre 2003. Parte Teórica. 6 puntos

1. A partir de una muestra de tamaño n proveniente de una distribución de Poisson de parámetro λ , obtener el estimador máximo verosímil para dicho parámetro.
2. A 1000 alumnos que comienzan una determinada carrera se les ha seguido la pista durante 10 años. Al cabo de dicho tiempo nos encontramos con que 92 fueron capaces de terminar la carrera en 5 años, 218 necesitaron 6 años, 275 lo hicieron en 7 años, 312 en 8 años y los 103 restantes necesitaron 9 o más años para finalizar sus estudios.

Nos preguntamos si podemos seguir manteniendo la hipótesis de que los datos provienen de una distribución de probabilidad polinomial con parámetros $p_1 = \frac{1}{10}$, $p_2 = \frac{2}{10}$, $p_3 = \frac{3}{10}$, $p_4 = \frac{3}{10}$ y $p_5 = \frac{1}{10}$. Efectuar el test con un nivel de significación de $\alpha = 0.05$.

3. Pon tres ejemplos de tu invención en los que tenga interés abordar los problemas con técnicas de clasificación supervisada. Especifica claramente el problema, así como las variables predictoras y la variable clase.
4. En la actualidad los partidos de baloncesto se desarrollan en cuatro tiempos de 10 minutos cada uno. Pretendemos desarrollar un clasificador para predecir si un equipo determinado va a ganar un partido concreto. Para ello nos vamos a fijar exclusivamente en los resultados de los dos primeros tiempos, más concretamente en si dicho equipo ha ganado o ha perdido cada uno de los dos primeros tiempos.

Hemos recopilado datos de un buen número de partidos y nos hemos percatado de que tan sólo el 40 por ciento de los equipos que pierden el primer tiempo es capaz de ganar el partido, mientras que si el equipo ganase el primer tiempo dicha probabilidad subiría hasta el 0.55. En relación con el segundo tiempo, constatamos que el 35 por ciento de los equipos que pierden dicho segundo tiempo son capaces de ganar el partido, mientras que para los equipos que ganan el segundo tiempo dicha proporción aumenta hasta el 60 por ciento.

(a) Con la información que disponemos ¿Cuál de los paradigmas de clasificación supervisada vistos durante el curso puede ser aplicado? Razona tu respuesta.

(b) Efectúa el pronóstico para un equipo que ha perdido el primer tiempo, pero ha ganado el segundo tiempo.

5. Tanto la edición de Wilson como la condensación de Hart son dos técnicas heurísticas cuya finalidad es reducir el tamaño del conjunto de casos a almacenar por el clasificador k -NN.

Diseña un algoritmo genético (representación de los individuos, operadores de cruce y mutación, y función de evaluación) con ese mismo objetivo.

6. Explica las diferencias entre los algoritmos de Forgy y McQueen. Pon un ejemplo en el que el resultado de aplicar ambos algoritmos no coincida.