

# Métodos Probabilísticos en Inteligencia Artificial

Junio 2001. Examen Teórico. 6 puntos

1. Se extrae una muestra aleatoria de tamaño  $n$  de una distribución normal,  $X$ , con parámetros  $\mu$  y  $\sigma$ , ambos desconocidos. Obtener los estimadores máximo verosímiles para dichos parámetros
2. Los 300 alumnos matriculados por vez primera durante este curso en la asignatura A han obtenido los siguientes resultados en su convocatoria de junio:
  - 55 no presentados
  - 160 suspensos
  - 60 aprobados
  - 20 notables
  - 5 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:  $p_1 = \frac{4}{20}, p_2 = \frac{10}{20}, p_3 = \frac{4}{20}, p_4 = \frac{1}{20}, p_5 = \frac{1}{20}$ .

3. Es conocido que la condensación de Clark para la selección de prototipos en el paradigma  $K$ -NN depende del orden en el que se encuentran almacenados los objetos. Diseñar un algoritmo genético (representación de individuos, función de evaluación, operadores de cruce y mutación) para la búsqueda del mejor orden con el que se deben de almacenar los objetos.
4. Partiendo del algoritmo de MacQueen, diseña un algoritmo de clasificación no supervisada particional donde el número de clusters,  $k$ , pueda ser adaptado a medida que se desarrolle el algoritmo.
5. Supongamos que se conoce que la probabilidad a priori de recidiva para un enfermo que sufre por vez primera un ataque de corazón es de 0.30.

Supongamos que en el modelo que tenemos construido para tratar de predecir la recidiva, tenemos dos variables predictoras que se suponen independientes dado el valor de la clase. Las dos variables predictivas son la *realización de ejercicio moderado de manera asidua* y la *edad del paciente*. Mientras que la primera es discreta con dos posibles valores (sí, no), la segunda es continua y está para cada valor de la clase distribuida normalmente.

Entre los que no tienen recidiva, el 80 por ciento efectúa ejercicio moderado de manera asidua, mientras que entre los que tienen recidiva dicha probabilidad se reduce al 10 por ciento.

La edad de los pacientes que no tienen recidiva sigue un modelo normal con esperanza matemática de 60 años y desviación típica de 10 años, mientras que en el caso de los pacientes con recidiva los parámetros del modelo normal son 75 años de esperanza matemática y 15 de desviación típica.

Efectuar el diagnóstico para un paciente de 58 años que no efectúa ejercicio moderado de manera asidua.

6. Explicar las diferencias y similitudes entre los siguientes métodos de validación de modelos clasificatorios: submuestreo aleatorio, validaciones cruzadas de  $k$  rodajas, método *bootstrapping*.