

APRENDIZAJE AUTOMATICO

Cuestionario-T2: 8 puntos

Fecha entrega: 26 Abril

Justificar la contestación en todos los apartados
Incluir los enunciados en la contestación

- 1) **(1.5 puntos)** Leer el apartado 4.4.2 del libro ISLR
 - a) Deducir las expresiones 4.13 y 4.14 a partir de la expresión 4.12. Justificar cada uno de los pasos dados.
 - b) Interpretar la relación que existe entre las expresiones 4.12 y 4.13 en el contexto de la clasificación de una observación.
 - c) Interpretar la relación que existe entre las expresiones 4.12 y 4.14 en el contexto de la clasificación de una observación.
- 2) **(2 puntos)** Supongamos k clases distintas representadas por distribuciones Normales univariantes de media μ_k y varianza σ_k^2 respectivamente. Deducir la regla de Bayes asociada a asignar una nueva observación a la clase más probable. ¿Es una función lineal o cuadrática? Justificar la respuesta
- 3) **(1 punto)**
 - a) En promedio, ¿qué fracción de gente con un cociente de probabilidades (odds) de 0.37 de no pagar el gasto de su tarjeta de crédito, realmente no pagará? Justificar la respuesta
 - b) Supongamos que un individuo tiene 16% de posibilidades de no pagar con su tarjeta de crédito. ¿Cuánto vale el cociente de probabilidades (odds) de que no pague? Justificar la respuesta
- 4) **(1.5 puntos)** Supongamos que tenemos datos de un grupo de estudiantes de un curso de AA. Las muestras se componen de X_1 =horas estudiadas, X_2 =puntuación de formación del estudiante (valor de 1-4), e Y = haber obtenido un sobresaliente . Ajustamos un modelo de regresión logística y estimamos como coeficientes $\beta_0 = -6$, $\beta_1 = 0.05$, $\beta_2 = 1$.
 - a) Estimar la probabilidad de que un estudiante que estudia 40 horas y su puntuación en formación es de 3.5, obtenga un sobresaliente .
 - b) ¿Cuántas horas necesita estudiar un estudiante para tener al menos un 50% de posibilidades de obtener un sobresaliente?
- 5) **(2 puntos)** Supongamos que deseamos predecir, a partir de X (el porcentaje de beneficio del último año) si una determinada inversión en bolsa dará dividendo en este año (Si o No). Analizamos un conjunto de compañías y descubrimos que el valor medio de X para las compañías que dieron dividendo en el pasado fue $X=10$, mientras que la media de aquellas que no lo dieron fue $X=0$. Además la varianza estimada de X para estos dos conjuntos de compañías fue $\sigma^2 = 36$. Finalmente, el 80% de las compañías dieron dividendos.
 - a) Suponiendo que X sigue una distribución Normal, predecir la probabilidad de que una compañía de dividendo este año dado que su porcentaje de beneficios del último año fue $X=4$

BONUS-1 (2.5 puntos) Suponga que tomamos un conjunto de datos y lo dividimos en dos partes iguales, una para entrenamiento y otra para test y aplicamos dos técnicas de clasificación distintas. Primero usamos Regresión Logística y obtenemos una tasa de error del 20% sobre los datos de entrenamiento y del 30% sobre los datos de test. A continuación usamos k-NN con $k=1$ y obtenemos una tasa promedio de error (promedio entre entrenamiento y test) del 18%. Basándose en estos resultados, ¿qué método deberíamos preferir usar para clasificar nuevas observaciones? Justificar adecuadamente la respuesta.