Cuestionario 2

Aprendizaje Automático

Grado en Ingeniería Informática

Granada, 2 de Mayo de 2015.

Datos del estudiante

Fernández Bosch, Pedro
76422233-H

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.

Utilizar el clasificador de Bayes, implica la asignación de una observación (k) en donde el resultado sea mayor.

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2_{\pi\sigma}}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)} = \frac{\pi_k e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}}{\sum_{l=1}^{K} \pi_l e^{\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}}$$

Como la función es estrictamente creciente, equivale a encontrar k en donde el resultado sea mayor.

$$\log p_k(x) = \log \pi_k \left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right) - log \sum_{l=1}^{K} \pi_l \, e^{\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

El último término es independiente de k, por lo tanto es posible limitar la búsqueda del valor k a donde el resultado sea mayor.

$$\log \pi_k - \left(\frac{1}{2\sigma^2}(x-\mu_k)^2\right) = \log \pi_k - \frac{1}{2\sigma^2}x^2 + \frac{\mu_k}{\sigma^2}x - \frac{\mu_k^2}{2\sigma^2}$$

El término en x^2 es independiente de k, por lo que se mantiene, para encontrar el valor de k en donde el resultado sea mayor.

$$\delta_k(x) = x.\frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \log(\pi_k)$$

Si por ejemplo, K = 2 y $\pi_1 = \pi_2$, entonces el clasificador de Bayes asigna una observación a la clase 1 si 2x $(\mu_1 - \mu_2) > \mu_1^2 - \mu_2^2$, y en caso contrario la asigna a la clase 2.

$$x = \frac{\mu_1^2 - \mu_2^2}{2(\mu_1 - \mu_2)} = \frac{\mu_1 + \mu_2}{2}$$

b) Interpretar la relación que existe entre las expresiones 4.12 y 4.13 en el contexto de la clasificación de una observación.

El clasificador de Bayes implica la asignación de una observación X = x a la clase más grande. Aplicando el logaritmo y reordenando los términos, se puede demostrar que esto es equivalente a la asignación de la observación a la clase para la que es más grande.

c) Interpretar la relación que existe entre las expresiones 4.12 y 4.14 en el contexto de la clasificación de una observación.

Si por ejemplo, K = 2 y $\pi_1 = \pi_2$, entonces el clasificador de Bayes asigna una observación a la clase 1 si 2x $(\mu_1 - \mu_2)$ > $\mu_1^2 - \mu_2^2$, y en caso contrario la asigna a la clase 2.

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

Se ha procedido de la misma manera que en el apartado anterior. Se pretende afirmar que la búsqueda de k para el cual $p_k(x)$ es el más grande, equivale a encontrar k para el que es más grande. Por lo tanto:

$$\log \pi_k - \left(\frac{1}{2\sigma^2}(x-\mu_k)^2\right) = \log \pi_k - \frac{1}{2\sigma^2}x^2 + \frac{\mu_k}{\sigma^2}x - \frac{\mu_k^2}{2\sigma^2} - log\pi_k$$

<u>Esta expresión no es lineal en x.</u>

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

Odds es una expresión numérica, que consta de un par de números utilizados tanto en juegos de azar como en estadísticas. En las estadísticas, odds pretende calcular la probabilidad de que un evento en particular ocurra. La cantidad de $\frac{p(X)}{1-p(X)}$ se llama odds.

$$\frac{p(X)}{1-p(X)} = 0{,}37$$

$$p(X) = 0{,}27$$

<u>Según el valor obtenido al despejar en la fórmula de odds, en promedio, el 27% de gente realmente no pagará.</u>

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 page? Justificar la respuesta

Para resolver este problema, se va a utilizar la misma ecuación del apartado anterior.

$$\frac{0{,}16}{1-0{,}16} = 0{,}19$$

<u>Por lo tanto, el cociente de probabilidades (odds) de que no page vale 0,19.</u>

4) (1.5 puntos) Supongamos que tenemos datos de un grupo de estudiantes de un curso de AA. Las muestras se componen de X1 =horas estudiadas, X2 =Formación del estudiante (1-4), e Y = haber obtenido un sobresaliente (Y>3). Ajustamos un modelo de regresión logística y estimamos como coeficientes β0 = −6, β1 = 0.05, β2 = 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 (Y>3).

Para resolver este problema se va a utilizar la función Logistica. En lugar de predecir Y, predecimos P(Y > 3).

$$p = P(Y > 3 \mid X) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2}} = \frac{e^{-6+0{,}05\cdot 40+1\cdot 3{,}5}}{1+e^{-6+0{,}05\cdot 40+1\cdot 3{,}5}} = 0{,}38$$

Por lo tanto, la probabilidad de que un estudiante que estudia 40 horas y su puntuación en formación es de 3.5, obtenga un sobresaliente es del 38%

b) ¿Cuantas horas necesita estudiar un estudiante para tener al menos un 50% de posibilidades de obtener un sobresaliente?

0.5 = exp(−2.5 + 0.05x)/(1 + exp(−2.5 + 0.05x)).

$$0{,}5 = \frac{e^{-6+0{,}05\cdot X_1+1\cdot 3{,}5}}{1+e^{-6+0{,}05\cdot X_1+1\cdot 3{,}5}} = \frac{e^{-2{,}5+0{,}05X_1}}{1+e^{-2{,}5+0{,}05X_1}}$$

$$X_1 = 50$$

Para tener al menos un 50% de posibilidades de obtener un sobresaliente, un estudiante necesita estudiar 50 horas.

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 σ2 = 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

Según el enunciado, tenemos que:
$x=4,\ \sigma^2=36,\ \mu_1=10,\ \mu_0=0,\ \pi_1=0.8,\ \pi_0=0.2,$

Para resolver, sólo hay que sustituir los parámetros y valores de x en la ecuación para $P_k(x)$:

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2_{\pi\sigma}}} exp\left(-\frac{1}{2\sigma^2}(x-\mu_l)^2\right)}$$

$$p_1(4) = \frac{0{,}8 \exp\left(-\left(\frac{1}{72}\right)(4-10)^2\right)}{0{,}8 \exp\left(-\left(\frac{1}{72}\right)(4-10)^2\right) + 0{,}2 \exp\left(-\left(\frac{1}{72}\right)(4-0)^2\right)} = 0{,}75$$

<u>Por lo tanto, se ha obtenido que la probabilidad de que una compañía de dividendo este año dado que su porcentaje de beneficios del último año fue X=4 es del 75%.</u>

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.

Con los datos que aporta el enunciado conocemos que:

Para Regresión Logistica:
- La tasa de error en train es del: 20%.
- La tasa de error en test es del: 30%.

Para k-NN con k=1:
- La tasa de error promedio de train y test es del 18%.

Con los datos que aporta el problema no es posible decidir cuál de los dos métodos es preferible.

A continuación se va a analizar un supuesto ejemplo en que la tasa de error en train podría ser sólo del 1% (posible sobreajuste) y por lo tanto, la tasa de error en train valdría el 35%. Esto significa que $\frac{3+35}{2}$= 18. Es decir, la tasa de error promedio de train y test en k-NN con k=1 sería del 18%, no obstante, la tasa de error en test es mayor en k-NN con k=1 que en Regresión Lineal.

La tasa de error en train no es tan relevante como la tasa de error en test. Resulta fundamental que la tasa de error en test sea lo mejor posible, ya que este dato va determinar como de bien se clasificarán nuevas observaciones.

<u>En este ejemplo sería aconsejable tomar Regresión Lineal para clasificar nuevas observaciones, pero en cualquier caso, la elección depende del error de test para k-NN=1.</u>