Mario López Díaz
Ignacio Gutiérrez Sánchez

Tarea 3

13.1 Aplicamos la fórmula de Bayes

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \longrightarrow p(class\,1|x) = \frac{p_1 p(x|class\,1)}{p_1 p(x|class\,1) + p_2 p(x|class\,2)}$$

Igualamos a 0,5 porque $p(class\,1|x) = 0{,}5$

$$\frac{p_1 p(x|class\,1)}{p_1 p(x|class\,1) + p_2 p(x|class\,2)} = 0{,}5 \longrightarrow p_1 p(class\,1|x) = p_2 p(x|class\,2)$$

Sustituimos las normales

$$p_1 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = p_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

Aplicamos logaritmos

Despejamos

$$\log p_1 - \frac{(x-\mu_1)^2}{2\sigma_1^2} = \log p_2 - \frac{(x-\mu_2)^2}{2\sigma_2^2} \longrightarrow \frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2} = \log \frac{p_2}{p_1}$$

$$\rightarrow x^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) + x\left(\frac{-2\mu_1}{\sigma_1^2} + \frac{2\mu_2}{\sigma_2^2}\right) + \left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2}\right) = 2\log\frac{p_2}{p_1}$$

Tiene la forma general de ecuación cuadrática
$Ax^2 + Bx + C = 0$

Al aplicar $x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$ nos damos cuenta que dependiendo del discriminante $B^2 - 4AC$ puede suceder que haya:

- Dos soluciones si el discriminante es positivo
- Una solución si el discriminante es igual a cero
- Ninguna solución real si el discriminante es negativo

Si obtenemos dos soluciones es razonable decir que hay dos puntos donde las probabilidad de ambas clases son iguales

Si obtenemos una solución la frontera de decisión sería un punto, esto es razonable si las varianzas σ_1^2 y σ_2^2 son iguales

Si no hay soluciones significa que las clases no pueden separarse bajo el modelo propuesto

13.2 Paso 1 Error bajo la regla de decisión Bayesiana

$$\text{Error Bay} = 1 - p(\text{class } c|x)$$ class c = clase correcta

Paso 2 Error bajo una regla de decisión aletoria

$$\text{Error aleatorio} = \sum_{j \neq K} q(\text{class } j|x)\, p(\text{class } K|x) = 1 - q(\text{class } j|x)\, p(\text{class } K|x)$$

Paso 3

EL error se minimiza cuando $q(\text{class } K|x) = 1$ para la clase con mayor probabilidad

Esto es la regla de decisión Bayesiana

13.3 • Utilidad esperada para $c_{pred} = 1$

$$0{,}7 \cdot 5 + 0{,}2 \cdot 0 + 0{,}1 \cdot (-3) = \boxed{3{,}2}$$

• Utilidad esperada para $c_{pred} = 2$

$$0{,}7 \cdot 3 + 0{,}2 \cdot 4 + 0{,}1 \cdot 0 = \boxed{2{,}9}$$

• Utilidad esperada para $c_{pred} = 3$

$$0{,}7 \cdot 1 + 0{,}2 \cdot (-2) + 0{,}1 \cdot 10 = \boxed{1{,}3}$$

La mejor decesión es para $\boxed{c_{pred} =}$ con 3,2

La mejor desisión es predecir la clase 1

13.5 Planteamos 4 cuestiones:

- Credibilidad: Aunque la red haya logrado 0 de error en todos los problemas conocidos esto no garantiza que funcione para problemas futuros
- Memorización de los conjuntos de prueba
- Complejidad del modelo: Una red con un billón de capas ocultas tiene un alto riesgo de overfitting

Conclusión

El hecho de que el modelo prediga pefectamente problemas conocidos no es suficiente para justificar su compra