Ciencia de Datos

Práctico N°2: Teoría Bayesiana

Problema 1: En el caso de dos categorías, según la regla de decisión de Bayes, el error condicional viene dado por la ecuación,

$$P(\text{error}|x) = \text{mín}[P(\omega_1|x), P(\omega_2|x)]\,,$$

donde ω_i denota los posibles estados del sistema y x es una variable aleatoria cuyo valor depende del estado del sistema. Incluso si las densidades a posteriori $P(\omega_i|x)$ son continuas, el error condicional casi siempre conduce a un integrando discontinuo en el calculo del error total

$$P(\text{error}) = \int P(\text{error}|x)\, p(x)\, dx\,.$$

a) Demostrar que para densidades arbitrarias, una cota superior para el error total resulta del hecho de que siempre se cumple que

$$P(\text{error}|x) \leq 2\, P(\omega_1|x)\, P(\omega_2|x)\,.$$

b) Demostrar que si en la expresión para $P(\text{error})$ se sustituye según $P(\text{error}|x) = \alpha\, P(\omega_1|x)\, P(\omega_2|x)$, con $\alpha < 2$, entonces no puede garantizarse que la integral sea una cota superior para el error.
c) Análogamente, demostrar que puede utilizarse $P(\text{error}|x) = P(\omega_1|x)\, P(\omega_2|x)$ para obtener una cota inferior para el error total.
d) Demostrar que si $P(\text{error}|x) = \beta\, P(\omega_1|x)\, P(\omega_2|x)$ con $\beta > 1$, entonces la integral puede no ser una cota inferior para el error.

Problema 2: Suponer dos variables aleatorias independientes idénticamente distribuidas con la densidad de Laplace,

$$p(x|\omega_i) \propto \exp\left(-\frac{|x - a_i|}{b_i}\right)\,, \text{ con } i = 1, 2 \text{ y } b_i > 0\,.$$

a) Escribir las expresiones analíticas normalizadas de $p(x|\omega_i)$.
b) Calcular el radio de verosimilitud como función de los parámetros.
c) Graficar el radio $p(x|\omega_1)/p(x|\omega_2)$ para el caso $a_1 = 0$, $b_1 = 1$, $a_2 = 1$ y $b_2 = 2$.

Problema 3: Considerar la siguiente regla de decisión para el problema unidimensional con dos categorías: Se decide por ω_1 si $x > \theta$ y en otro caso se decide por ω_2.
a) Demostrar que la probabilidad de error para esta regla viene dada por

$$P(error) = \int P(error|x)p(x)dx = P(\omega_1)\int_{-\infty}^{\theta} p(x|\omega_1)dx + P(\omega_2)\int_{\theta}^{\infty} p(x|\omega_2)dx\,.$$

b) Demostrar que una condición necesaria para minimizar el error es $p(\theta|\omega_1)\, p(\omega_1) = p(\theta|\omega_2)\, p(\omega_2)$.
c) ¿Define esta ecuación un valor de θ único?
d) Estudiar como ejemplo el caso en el que la variable X condicional a ω_i tiene distribución normal con media μ_i y desvío σ_i; es decir, $P(X|\omega_i) \sim N(\mu_i, \sigma_i)$.

Problema 4: Suponer que se sustituye la función de decisión determinista $\alpha(x)$ por la regla aleatoria dada por la probabilidad $P(\alpha_i|x)$ de tomar la decisión α_i dado que se observo x.

a) Mostrar que el riesgo resultante viene dado por,

$$R = \int \left(\sum_{i=1}^{a} R(\alpha_i|x)\, P(\alpha_i|x) \right) p(x) dx\,.$$

b) Demostrar además que R se minimiza para $P(\alpha_i|x) = 1$ para la acción α_i asociada con el riesgo condicional mínimo $R(\alpha_i|x)$, lo que demuestra que no obtenemos ningún beneficio haciendo aleatoria la regla de decisión.

Problema 5: En muchos problemas de clasificación multicategoría: ω_i con $i = 1, \ldots, c$, es conveniente trabajar con una función de pérdida pesada. Por ejemplo, puede ocurrir que se rechace un patrón o estado del sistema si este resulta irreconocible,

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & \text{si } i = j,\ i, j = 1, 2, \ldots, c\,, \\ \lambda_r & \text{si } i = c + 1\,, \\ \lambda_s & \text{en otro caso.} \end{cases}$$

donde λ_r es la pérdida sufrida por la elección de rechazo, λ_s es la pérdida incurrida por cometer un error.
Mostrar que el riesgo mínimo se obtiene si decidimos α_i si $P(\omega_i|x) \geq P(\omega_j|x)$ para todo j, y si $P(\omega_i|x) \geq 1 - \dfrac{\lambda_r}{\lambda_s}$, caso contrario, rechazar. ¿Que sucede si $\lambda_r = 0$? ¿Que sucede si $\lambda_r > \lambda_s$?

Problema 6: Retomar el problema de clasificación con la opción de rechazo del problema anterior.
a) Demostrar que las siguientes funciones discriminantes son óptimas para este tipo de problemas:

$$g_i(x) = \begin{cases} p(x|\omega_i)\, P(\omega_i) & \text{si } i = 1, 2, \ldots, c\,, \\ \dfrac{\lambda_s - \lambda_r}{\lambda_s} \displaystyle\sum_{j=1}^{c} p(x|\omega_j)\, P(\omega_j) & \text{si } i = c + 1\,. \end{cases}$$

b) Graficar la función discriminante y las regiones de decisión para el caso del problema unidimensional ($x \in \mathcal{R}$) con dos clases usando los valores

$$p(x|\omega_1) \sim \mathcal{N}(1, 1),\ p(x|\omega_2) \sim \mathcal{N}(-1, 1),\ P(\omega_1) = P(\omega_2)\,,\ \frac{\lambda_r}{\lambda_s} = \frac{1}{4}\,.$$

c) Describir cualitativamente lo que sucede cuando $\dfrac{\lambda_r}{\lambda_s}$ se incrementa desde 0 a 1.
d) Considerar nuevamente este problema, ahora en el caso particular

$$p(x|\omega_1) \sim \mathcal{N}(1, 1), p(x|\omega_2) \sim \mathcal{N}\left(0, \frac{1}{4}\right),\ P(\omega_1) = \frac{1}{3}\,, P(\omega_2) = \frac{2}{3}\,,\ \frac{\lambda_r}{\lambda_s} = \frac{1}{2}\,.$$

Problema 7: Estudiar la implementación del análisis de discriminante lineal provista por scikit-learn para generar muestras aleatorias de acuerdo a una distribución normal bivariada y calcular la función discriminante para una distribución normal dada y probabilidades a priori $P(\omega_i)$.
a) Simular dos variables normales (X_1, X_2) con $\Sigma = C^T.C$, y $C = \begin{pmatrix} 0 & -0{,}23 \\ 0{,}83 & 0{,}23 \end{pmatrix}$ y vectores de medias $\mu_1 = (0, 0)$ y $\mu_2 = (1, 1)$, respectivamente.
b) Suponer que las probabilidades a priori de las dos categorías son iguales ($P(\omega_1) = P(\omega_2)$), e implementar un clasificador para dos categorías utilizando sólo el valor de característica X_1

especificada en el inciso anterior. El código resultante debe poder clasificar una nueva muestra basado en esta información.

Tener presente que para el diseño del clasificador se estimará la media y varianza a partir de los datos de cada una de las muestras. Si para la muestra i la media y varianza son μ_i y σ_i^2 respectivamente, se clasificará un valor x en la muestra 1 si

$$\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(x-\mu_1)^2/2\sigma_1^2} P(\omega_1) > \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(x-\mu_2)^2/2\sigma_2^2} P(\omega_2) \,.$$

Tomando logaritmo, y eliminando las probabilidades a priori $P(\omega_i)$ por ser iguales esto es equivalente a decidir por la clase 1 si

$$-\frac{1}{2}\ln(2\pi) - \ln\sigma_1 - \frac{(x-\mu_1)^2}{2\sigma_1^2} > -\frac{1}{2}\ln(2\pi) - \ln\sigma_2 - \frac{(x-\mu_2)^2}{2\sigma_2^2}\,;$$

es decir, si

$$\ln\sigma_1 + \frac{(x-\mu_1)^2}{2\sigma_1^2} < \ln\sigma_2 + \frac{(x-\mu_2)^2}{2\sigma_2^2}\,.$$

c) Determinar el error de entrenamiento empírico en la clasificación muestras; esto es, el porcentaje de puntos mal clasificados, dividiendo aleatoriamente el número de muestras $n = 100$, en 80 % entrenamiento y en 20 % test. Repetir incrementando los valores de n, desde 100 a 10000 en pasos de 100 y graficar el error empírico obtenido.

d) Utilizar la cota de Bhattacharyya para acotar el error que obtendrán los nuevos patrones obtenidos muestreando las distribuciones.

e) Repetir todo lo anterior, pero ahora utilice las dos características, X_1 y X_2.

f) Analizar resultados. ¿Es siempre posible para un conjunto finito de datos que el error empírico resulte mayor al aumentar la dimensión de los datos?

Problema 8: La distribución de Poisson para una variable entera no negativa $x = 0, 1, \ldots$ y parámetro real λ viene dada por

$$P(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$$

Considerar el problema de clasificación con dos categorías igualmente probables $P(\omega_1) = P(\omega_2)$ y condicionales con distribuciones de Poisson con diferentes parámetros $\lambda_1 > \lambda_2$.

a) Especificar regla de clasificación de Bayes.

b) ¿Cuál es la tasa del error de Bayes?

c) Escribir función discriminante, y determinar qué valores debe tener de entrada para clasificar un nuevo dato.

d) Simular una muestra aleatoria de tamaño 100 con distribuciones de Poisson con $\lambda_1 = 1{,}8$, $\lambda_2 = 0{,}4$, considerando igual probabilidad a priori. Usar la función de pérdida cero uno y clasificar la muestra acorde a esta función. Estimar el error cometido en la muestra, y compararlo con el error de Bayes calculado.