PRÁCTICA I

TEORÍA DE LA DECISIÓN BAYESIANA

Ejercicios Teóricos

Ejercicio 1. En el caso de dos categorías, en la regla de decisión de Bayes el error condicional está dado por la ecuación (7). Incluso si las densidades a posteriori son continuas, esta forma de la condicional de errores casi siempre conduce a un integrando discontinuo en el cálculo del error total en la ecuación (5).

a) Demostrar que para densidades arbitrarias, podemos usar la ecuación (5)

$$P(error|x) = 2P(\omega_1|x)P(\omega_2|x)$$

y obtener una cota superior para el error total.

b) Demostrar que si en la ecuación (5), utilizamos $P(error|x) = \alpha P(\omega_1|x)P(\omega_2|x)$, con $\alpha < 2$, entonces no podemos garantizar que la integral dé una cota superior para el error.

c) Análogamente, demostrar que podemos utilizar

$$P(error|x) = P(\omega_1|x)P(\omega_2|x)$$

y obtener una cota inferior para el error total.

d) Demostrar que si

$$P(error|x) = \beta P(\omega_1|x)P(\omega_2|x)$$

con $\beta > 1$, entonces no podemos garantizar que la integral dé una cota inferior para el error.

Ejercicio 2. Si las distribuciones condicionales para las dos categorías en el problema unidimensional son distribuciones de Cauchy tal como se describe en el problema 6.

a) por integración explícita, compruebe que las distribuciones están normalizadas.

b) Suponiendo que $P(\omega_1) = P(\omega_2)$, muestre que $P(\omega_1|x) = P(\omega_2|x)$ si $x = \frac{a_1+a_2}{2}$, es decir, el límite inferior del error de decisión es un punto medio entre los picos de las dos distribuciones, independientemente de b.

c) Graficar $P(\omega_1|x)$ para el caso de $a_1 = 3$, $a_2 = 5$ y $b = 1$.

d) ¿Cómo se comportan $P(\omega_1|x)$ y $P(\omega_2|x)$ cuando $x \to \infty$ y cuando $x \to -\infty$? Explicar.

Ejercicio 3. Use la densidad condicional dada en el problema 6, y asuma la igualdad entre las probabilidades a priori para las categorías.

a) Demostrar que la probabilidad de error mínimo está dado por

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|$$

b) Graficar esto como una función de $\dfrac{a_2 - a_1}{b}$.

c) ¿Cuál es el máximo de $P(error)$ y en qué condiciones ocurre esto? Explicar.

Ejercicio 4. Responda con Verdadero o Falso, y justifique:

a) En un problema unidimensional con dos categorías y x continua, una transformación monótona de x no modifica la tasa de error de Bayes.

b) En un problema bidimensional con dos categorías y x continua, una transformación monótona en x_1 y en x_2 no modifica la tasa de error de Bayes.

Ejercicio 5. Supongamos que se sustituye la función de decisión determinista $\alpha(x)$ por la regla aleatorizada, a saber, la probabilidad $P(\alpha_i|x)$ de tomar la decisión α_i dado que se observó 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) Además, demostrar 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 mejor regla de decisión.

c) ¿Podemos beneficiarnos aleatorizando una regla subóptima? Explicar.

Ejercicio 6. En muchos problemas de clasificación de patrones se tiene la opción de asignar un patrón a una de c clases, o de clasificarlo como irreconocible. Si el costo de hacer esto último no es demasiado alto, puede ser una acción deseable. Sea

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

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

Ejercicio 7. Considere el problema de clasificación con la opción irreconocible.

a) Utilice los resultados del ejercicio 6 para 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) Grafique esta función discriminante y las regiones de decisión para el caso del problema unidimensional con dos clases, teniendo

- $p(x|\omega_1) \sim \mathcal{N}(1, 1)$,
- $p(x|\omega_2) \sim \mathcal{N}(-1, 1)$,

- $P(\omega_1) = P(\omega_2) = \frac{1}{2}$, y
- $\frac{\lambda_r}{\lambda_s} = \frac{1}{4}$

c) Describa cualitativamente lo que sucede cuando $\frac{\lambda_r}{\lambda_s}$ se incrementa de 0 a 1.

d) Repita el procedimiento para el caso

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

Ejercicio 8. Considerar la frontera de decisión de Bayes para el caso de clasificación en dos categorías con d dimensiones.

a) Demostrar que para cualquier hipercubo arbitrario de d dimensiones, existen distribuciones normales tales que $p(x|\omega_i) \sim \mathcal{N}(\mu_i, \Sigma_i)$ y distribuciones apriori $P(\omega_i)$, $i = 1,2$, que poseen a este hipercubo como frontera de decisión de Bayes.

b) ¿Se mantiene esto si las apriori se mantienen fijas y no nulas, por ejemplo, $P(\omega_1) = P(\omega_2) = \frac{1}{2}$?

Ejercicio 9. Considere el problema de clasificación bidimensional en dos categorías con $p(x|\omega_1) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $p(x|\omega_2) \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{I}\right)$, y $P(\omega_1) = P(\omega_2) = \frac{1}{2}$

a) Calcular la cota de decisión de Bayes.

b) Calcular la cota el error de Bhattacharyya.

c) Repita los pasos anteriores con las mismas probabilidades apriori, pero con $p(x|\omega_1) \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix}\right)$ y $p(x|\omega_2) \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\right)$

Ejercicio 10. Supongamos que en el marco de detección de una señal tenemos dos gaussianas, pero con diferentes varianzas (cf., figura 2.20.), es decir, $p(x|\omega_1) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ y $p(x|\omega_2) \sim \mathcal{N}(\mu_2, \sigma_2^2)$, con $\mu_1 > \mu_2$ y $\sigma_1^2 \neq \sigma_2^2$. En este caso la curva ROC resultante no sería simétrica.

a) Supongamos que en este caso asimétrico se modifica la definición de discriminabilidad por

$$d'_a = \frac{|\mu_2 - \mu_1|}{\sqrt{\sigma_1 \sigma_2}}$$

Mostrar con un contraejemplo no trivial o análisis que no se puede determinar d'_a basándonos únicamente en un solo par de casos y falsas tasas de alarma.

b) Exponer y explicar todos los valores patológicos para los cuales esta fórmula no da un valor significativo para d'_a.

c) Grafique las curvas ROC para el caso $p(x|\omega_1) \sim \mathcal{N}(0,1)$ y $p(x|\omega_2) \sim \mathcal{N}(1,2)$.

Ejercicio 11. Suponiendo que $\lambda_{21} > \lambda_{11}$ y $\lambda_{12} > \lambda_{22}$, mostrar que la función discriminante de riesgo mínimo general para el caso binario independiente descripto en la Sección 2.9.1 está dado por

$$g(x) = w^t x + \omega_0$$

, donde w no cambia, y

$$\omega_0 = \sum_{i=1}^{d} \ln \frac{1-p_i}{1-q_i} + \ln \frac{P(\omega_1)}{P(\omega_2)} + \ln \frac{\lambda_{21}-\lambda_{11}}{\lambda_{12}-\lambda_{22}}$$

Ejercicio 12. La distribución de Poisson para una variable discreta $x = 0,1,2,\ldots$ y parámetro real λ es

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

a) Demostrar que la media de esta distribución es $E[x] = \lambda$.

b) Demuestre que la varianza de esta distribución es $Var[x] = \lambda$.

c) La *moda* de una distribución es el valor de x que tiene la mayor probabilidad. Demostrar que la moda de una distribución de Poisson es el mayor entero que no exceda λ, es decir, la moda es $\lfloor \lambda \rfloor$. (Si λ es un número entero, tanto λ como $\lambda - 1$ son modas de la distribución).

d) Considere el problema de clasificación en dos categorías igualmente probables con distribuciones de Poisson, pero con diferentes parámetros $\lambda_1 > \lambda_2$. ¿Cuál es la regla de clasificación de Bayes?

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

Ejercicios Computacionales

Ejercicio 13. Es posible que necesite del siguiente procedimiento en los siguientes ejercicios.

a) Escriba un procedimiento para generar muestras aleatorias de acuerdo a una distribución normal $\mathcal{N}_d(\mu, \Sigma)$.

b) Escriba un procedimiento que calcule la función discriminante (de la forma dada en La ecuación 47) para una distribución normal dada y probabilidad a priori $P(\omega_i)$.

c) Escriba un procedimiento que calcule la distancia euclídea entre dos puntos arbitrarios.

d) Escriba un procedimiento que calcule la distancia de Mahalanobis entre la media μ y un punto arbitrario x, dada la matriz de covarianza Σ.

Ejercicio 14. Utilice su clasificador del problema 13 para clasificar las 10 muestras de la tabla del libro de la siguiente manera. Suponga que las distribuciones subyacentes son normales.

a) Supongamos que las probabilidades a priori de las dos primeras categorías son iguales ($P(\omega_1) = P(\omega_2) = \frac{1}{2}$ y $P(\omega_3) = 0$) y el diseño de un separador para estas dos categorías utilizando sólo el valor de característica x_1.

b) Determinar el error de entrenamiento empírico en sus muestras, es decir, el porcentaje de puntos mal clasificados.

c) Utilice la cota de Bhattacharyya para acotar el error que obtendrá en los patrones nuevos, elaborados a partir de las distribuciones.

d) Repita todo lo anterior, pero ahora utilice dos características, x_1 y x_2.

e) Repita usando las tres características.

f) Analice sus resultados. En particular, ¿es siempre posible para un conjunto finito de datos que el error empírico pueda ser mayor al aumentar la dimensión de los datos?.

Ejercicio 15. Considerar las tres categorías del ejercicio anterior, y asuma que $P(\omega_i) = \frac{1}{3}$.

a) ¿Cuál es la distancia de Mahalanobis entre cada uno de los puntos $(1,2,1)^t, (5,3,2)^t, (0,0,0)^t, (1,0,0)^t$ y las medias de cada una de las clases.

b) Clasificar estos puntos.

c) Suponga que $P(\omega_1) = 0.8$ y $P(\omega_2) = P(\omega_3) = 0.1$, y clasifique nuevamente los puntos.

Ejercicio 16. Explora cómo hace o no el error empírico para acercarse a la cota de Bhattacharyya de la siguiente manera:

a) Escriba un procedimiento que genere muestras de una distribución normal $\mathcal{N}_d(\mu, \Sigma)$.

b) Considere $p(x|\omega_1) \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{I}\right)$, $p(x|\omega_2) \sim \mathcal{N}\left(\begin{pmatrix} -1 \\ 0 \end{pmatrix}, \mathbf{I}\right)$, con $P(\omega_1) = P(\omega_2) = \frac{1}{2}$. Por inspección establezca el límite de decisión de Bayes.

c) Generar $n = 100$ puntos (50 de ω_1 y 50 de ω_2) y calcular el error empírico.

d) Repita el procedimiento aumentando los valores de n, $100 \leq n \leq 1000$, en pasos de 100 y grafique los errores empíricos.

e) Analice sus resultados. En particular, ¿es siempre posible que el error empírico sea mayor que la cota de Bhattacharyya o Chernoff?.