Universidad Autónoma de Zacatecas
"Francisco García Salinas"
Maestría en Ciencias del Procesamiento de la Información

Teoría de decisión bayesiana

Porfirio Ángel Díaz Sánchez

24 de abril de 2021

Contenido

Contenido

Introducción

- La teoría de decisión bayesiana es un enfoque para la clasificación de patrones basado en la cuantificación de los *tradeoffs* entre varias decisiones de clasificación por medio de la probabilidad y costos de las mismas.
- Asume que el problema de la decisión se determina en términos probabilísticos y que se conocen todos los valores de probabilidad relevantes.

Contenido

Terminología

- Estado de naturaleza ω (variable aleatoria)
  - En el caso de peces, ω_1 para lubina, ω_2 para salmón.
- Probabilidades $P(\omega_1)$ y $P(\omega_2)$ (a priori)
  - Conocimiento a priori de la probabilidad de pescar una lubina o un salmón.
- (PDF) Función de densidad de probabilidad $P(x)$ (evidencia)
  - Frecuencia con la que se medirá un patrón con el valor x para la característica.

Terminología (2)

- Densidad de la probabilidad condicional $P(x|\omega_j)$ (likelihood)
  - Frecuencia con la que se medirá un patrón con el valor x para la característica dado que el patrón pertenece a la clase ω_j.

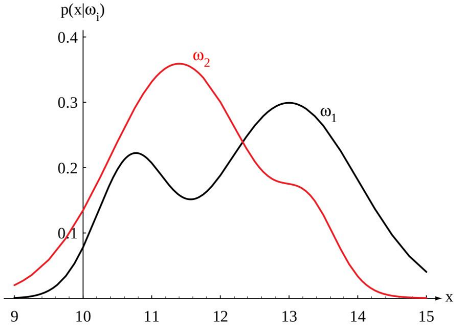


Figure 2.1: Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the length of a fish, the two curves might describe the difference in length of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0.

Terminología (3)

- Probabilidad condicional $P(\omega_j|x)$ (posterior)
  - Probabilidad de que el pescado pertenezca a la clase ω_j dada la medida x.

Contenido

Reglas de decisión usando probabilidades a priori

Decidir ω_1 si $P(\omega_1) > P(\omega_2)$; si no decidir ω_2

$$P(\text{error}) = \begin{cases} P(\omega_1) & \text{si } \omega_2, \\ P(\omega_2) & \text{si } \omega_1 \end{cases}$$

o

$$P(\text{error}) = min[P(\omega_1), P(\omega_2)]$$

- Favorece a la clase más probable.
- Toma la misma decisión todo el tiempo.

Contenido

Reglas de decisión usando probabilidades condicionales

Usando el teorema de bayes, la probabilidad posterior de la categoría ω_j dada la medición x se da por:

$$\boxed{P(\omega_j|x) = \frac{P(x|\omega_j)P(\omega_j)}{P(x)} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}}$$

Donde $P(x) = \sum_{j=1}^{2} P(x|\omega_j)P(\omega_j)$

Reglas de decisión usando probabilidades condicionales (2)

$$\text{Decidir } \omega_1 \text{ si } P(\omega_1|x) > P(\omega_2|x)\text{; si no decidir } \omega_2$$

o

$$\text{Decidir } \omega_1 \text{ si } P(x|\omega_1)P(\omega_1) > P(x|\omega_2)P(\omega_2)\text{; si no decidir } \omega_2$$

Reglas de decisión usando probabilidades condicionales (3)

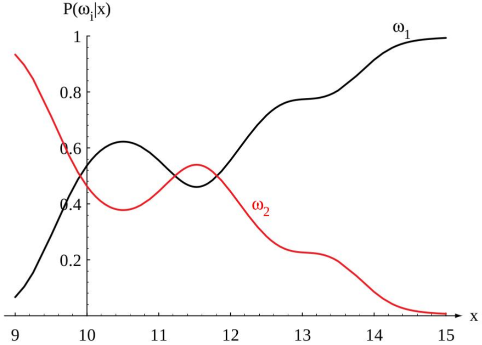


Figure 2.2: Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0.

Reglas de decisión usando probabilidades condicionales (4)

Probabilidad de error

$$P(\text{error}|x) = \begin{cases} P(\omega_1|x) & \text{si } \omega_2, \\ P(\omega_2|x) & \text{si } \omega_1 \end{cases}$$

$$o$$

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

Probabilidad de error promedio

$$P(\text{error}) = \int_{-\infty}^{\infty} P(\text{error, x})dx = \int_{-\infty}^{\infty} P(\text{error}|x)P(x)dx$$

Reglas de decisión usando probabilidades condicionales (5)

¿De donde vienen las probabilidades?

1. Enfoque de frecuencia relativa (objetivo)
   - Las probabilidades vienen de experimentos.
2. Enfoque bayesiano (subjetivo)
   - Las probabilidades pueden reflejar grados de creencia y pueden basarse en la opinión.

Contenido

Ejemplo

Enfoque objetivo: Clasificar de acuerdo a si su costo es mayor o menor a \$50K.

- **Clases:** C_1 si precio $>$ \$50K , C_2 si precio $<=$ \$50K
- **Características:** x, la altura del auto.

1. Por medio del teorema de Bayes se calculan las probabilidades posteriores:

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{P(x)}$$

2. Se requieren estimar $P(x|C_1)$, $P(x|C_2)$, $P(C_1)$ y $P(C_2)$

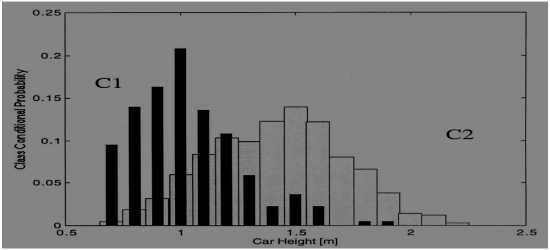
Ejemplo (2)

3. **Recolectar datos:** Preguntar a los conductores sobre el costo del auto y medir altura.
4. **Determinar probabilidades a priori:** $C_1 = 221,\ C_2 = 988$

   $P(C_1) = \frac{221}{1209} = 0.183,\ P(C_2) = \frac{988}{1209} = 0.817$

5. **Determinar probabilidades condicionales de clase (likelihood):**
   - Discretizar altura del auto en bins y usar histograma normalizado

$P(x|C_i)$



Ejemplo (3)

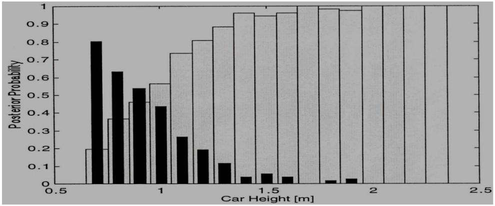
6. **Calcular la probabilidad posterior de cada bin:**

$$P(C_1|x=1.0) = \frac{P(x=1.0|C_1)P(C_1)}{P(x=1.0|C_1)P(C_1) + P(x=1.0|C_2)P(C_2)}$$

$$= \frac{(0.2081)(0.183)}{(0.2081)(0.183) + (0.0597)(0.817)}$$

$$= \mathbf{0.438}$$

$P(C_i|x)$



Contenido

Teoría general

- Usa más de una característica.
- Permite tener más de dos categorías.
- Posibilita realizar más acciones además de clasificar la entrada en una de las posibles categorías, por ejemplo, *rejection*.
- Emplea una función de error más general (función de riesgo), asociando un costo (función de pérdida) a cada error (acción incorrecta).

Contenido

Terminología

- Las características forman al vector $x \in R^d$ (espacio euclidiano d-dimensional), llamado espacio de características.
- Un conjunto finito de c categorías ω_1, ω_2, ..., ω_c.
- Regla de bayes, en notación de vector:

$$P(\omega_j|x) = \frac{P(x|\omega_j)P(\omega_j)}{P(x)}$$

$$\text{donde } P(x) = \sum_{j=1}^{c} P(x|\omega_j)P(\omega_j)$$

- Un conjunto finito de i acciones α_1, α_2, ..., α_i
- Una función de pérdida $\lambda(\alpha_i|\omega_j)$
  - El costo asociado de tomar la acción α_i cuando la clasificación de categoría correcta es ω_j.

Contenido

Riesgo condicional (pérdida esperada)

- Suponer que se observa x y se toma la acción $\boldsymbol{\alpha}_i$.
- Suponer que el costo asociado al tomar la acción $\boldsymbol{\alpha}_i$ con $\boldsymbol{\omega}_j$ siendo la categoría correcta es $\boldsymbol{\lambda}(\boldsymbol{\alpha}_i|\boldsymbol{\omega}_j)$.
- El riesgo condicional (o pérdida esperada) tomando la acción $\boldsymbol{\alpha}_i$ es:

$$R(\boldsymbol{\alpha}_i|x) = \sum_{j=1}^{c} \boldsymbol{\lambda}(\boldsymbol{\alpha}_i|\boldsymbol{\omega}_j) P(\boldsymbol{\omega}_j|x)$$

Contenido

Riesgo general

- Suponer $\boldsymbol{\alpha}(x)$ es una regla de decisión general que determina cuál acción $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, ..., $\boldsymbol{\alpha}_i$ tomar para cada x. El riesgo general se define como:

$$R = \int R(\boldsymbol{\alpha}(x)|x)P(x)dx$$

- La regla de decisión óptima es la regla de bayes, la cual minimiza a R así:
  - Calculando $R(\boldsymbol{\alpha}_i|x)$ para cada $\boldsymbol{\alpha}_i$ dada una x.
  - Escogiendo la acción $\boldsymbol{\alpha}_i$ con el mínimo $R(\boldsymbol{\alpha}_i|x)$.
- El riesgo mínimo resultante es llamado riesgo de bayes (R^*), y es el mejor rendimiento que se puede lograr.

Contenido

Ejemplo: Clasificación de dos categorías

- Se define que
  - Si $\boldsymbol{\alpha}_1$, se decide $\boldsymbol{\omega}_1$
  - Si $\boldsymbol{\alpha}_2$, se decide $\boldsymbol{\omega}_2$
  - $\boldsymbol{\lambda}_{ij} = \boldsymbol{\lambda}(\boldsymbol{\alpha}_i|\omega_j)$

- El riesgo condicional es:

$$R(\boldsymbol{\alpha}_i|x) = \sum_{j=1}^{c} \boldsymbol{\lambda}(\boldsymbol{\alpha}_i|\omega_j)P(\omega_j|x)$$

$$R(\boldsymbol{\alpha}_1|x) = \boldsymbol{\lambda}_{11}P(\omega_1|x) + \boldsymbol{\lambda}_{12}P(\omega_2|x)$$

$$R(\boldsymbol{\alpha}_2|x) = \boldsymbol{\lambda}_{21}P(\omega_1|x) + \boldsymbol{\lambda}_{22}P(\omega_2|x)$$

Ejemplo: Clasificación de dos categorías (2)

- Regla de decisión mínima:

$$\textbf{Decidir } \omega_1 \text{ si } R(\boldsymbol{\alpha}_1|x) < R(\boldsymbol{\alpha}_2|x)\text{; si no, } \boldsymbol{\omega}_2$$

o

$$\textbf{Decidir } \omega_1 \text{ si } (\boldsymbol{\lambda}_{21} - \boldsymbol{\lambda}_{11})P(\omega_1|x) > (\boldsymbol{\lambda}_{12} - \boldsymbol{\lambda}_{22})P(\omega_2|x)\text{; si no, } \omega_2$$

o (usando likelihood ratio)

$$\textbf{Decidir } \omega_1 \text{ si } \frac{P(x|\omega_1)}{P(x|\omega_2)} > \frac{(\boldsymbol{\lambda}_{12} - \boldsymbol{\lambda}_{22})}{(\boldsymbol{\lambda}_{21} - \boldsymbol{\lambda}_{11})} \frac{P(\omega_2)}{P(\omega_1)}\text{; si no, } \omega_2$$

La regla de bayes puede ser interpretada como la elección de ω_1 si el *likelihood ratio* excede el valor de *threshold* (umbral), que es independiente de la observación x.

Contenido

Caso especial: Función de pérdida Zero-One

- Asignar la misma pérdida a todos los errores:

$$\boldsymbol{\lambda}(\boldsymbol{\alpha}_i|\boldsymbol{\omega}_j) = \begin{cases} 0 & \text{si } i = j, \\ 1 & \text{si } i \neq j, \end{cases}$$

- El riesgo condicional corresponde a esta función de pérdida:

$$R(\boldsymbol{\alpha}_i|x) = \sum_{j=1}^{c} \boldsymbol{\lambda}(\boldsymbol{\alpha}_i|\boldsymbol{\omega}_j) P(\boldsymbol{\omega}_j|x) = \sum_{i \neq j} P(\boldsymbol{\omega}_j|x) = 1 - P(\boldsymbol{\omega}_j|x)$$

Caso especial: Función de pérdida Zero-One (2)

- Las reglas de decisión son:

$$\textbf{Decidir } \omega_1 \text{ si } R(\alpha_1|x) < R(\alpha_2|x)\text{; si no, } \omega_2$$

$$\text{o}$$

$$\textbf{Decidir } \omega_1 \text{ si } 1 - P(\omega_1|x) < 1 - P(\omega_2|x)\text{; si no, } \omega_2$$

$$\text{o}$$

$$\textbf{Decidir } \omega_1 \text{ si } P(\omega_1|x) > P(\omega_2|x)\text{; si no, } \omega_2$$

- En este caso, el riesgo general es el error de probabilidad promedio.