Teoría de Decisión Bayesiana

BAYESIAN DECISION THEORY

VARIABLE CONTINUA

Reglas de Decisión

- Observando la cinta de traslado es difícil predecir que pez va pasar primero
  - La secuencia de peces parece ser aleatoria.
- Terminología de teoría de decisión:
  - la naturaleza presenta dos estados
  - Cada estado es impredecible, por lo cual debe ser descrito en forma probabilística
- Estado verdadero $\boldsymbol{\omega}$ es una variable aleatoria
  - Dos estados verdaderos posibles:

$$\omega_1 = \text{Salmon}\ ,\ \omega_2 = \text{Mero}$$

Reglas de Decisión

- Las probabilidades de los estados reflejan el conocimiento previo sobre las características de los peces.
- Probabilidades a priori
  - Pesca del Salmon $P(\omega_1)$
  - Pesca del Mero $P(\omega_2)$

$$P(\omega_1) + P(\omega_2) = 1$$

- Por ejemplo si se suelen obtener tantos salmones como meros en la región, se puede decir que igualmente probable pescar un salmón que un mero si uno tira una línea.
  - $P(\omega_1) = P(\omega_2)$ *equiprobables*

Regla de Decisión solo con información a priori

- Decide ω_1 si $P(\omega_1) > P(\omega_2)$
- en otro caso decide ω_2

- Tiene sentido si solo juzgo el próximo pez.
- Error? Me equivoco con toda la clase menos probable!!!
- Sin embargo si no sé otra cosa, es la regla con menor error

Regla de Decisión con variable X

Por eso es que necesitamos diseñar y calcular caracterizaciones discriminatorias de los estados

- X es la luminosidad de las escamas del pez
- $p_X(x \mid \omega_1)$ y $p_X(x \mid \omega_2)$ describen la diferencia en luminosidad de las escamas entre poblaciones de Mero y Salmón
- $P(\omega_1)$ y $P(\omega_2)$ son las probabilidades a priori de la pesca
- Evidencia es la densidad de la variable X

$$\boxed{p_X(x) = \sum_{j=1}^{j=2} p(x \mid \omega_j) P(\omega_j)}$$

Regla de Decisión con variable X

Información condicional a la clase concentrada en una variable aleatoria X

- X es una variable aleatoria
- $p(x \mid \omega_1)$ y $p(x \mid \omega_2)$ probabilidades condicionadas a los estados
- $P(\omega_1)$ y $P(\omega_2)$ son las probabilidades a priori
- Evidencia es la densidad de la variable X

$$\boxed{p_X(x) = \sum_{j=1}^{j=2} p(x \mid \omega_j) P(\omega_j)}$$

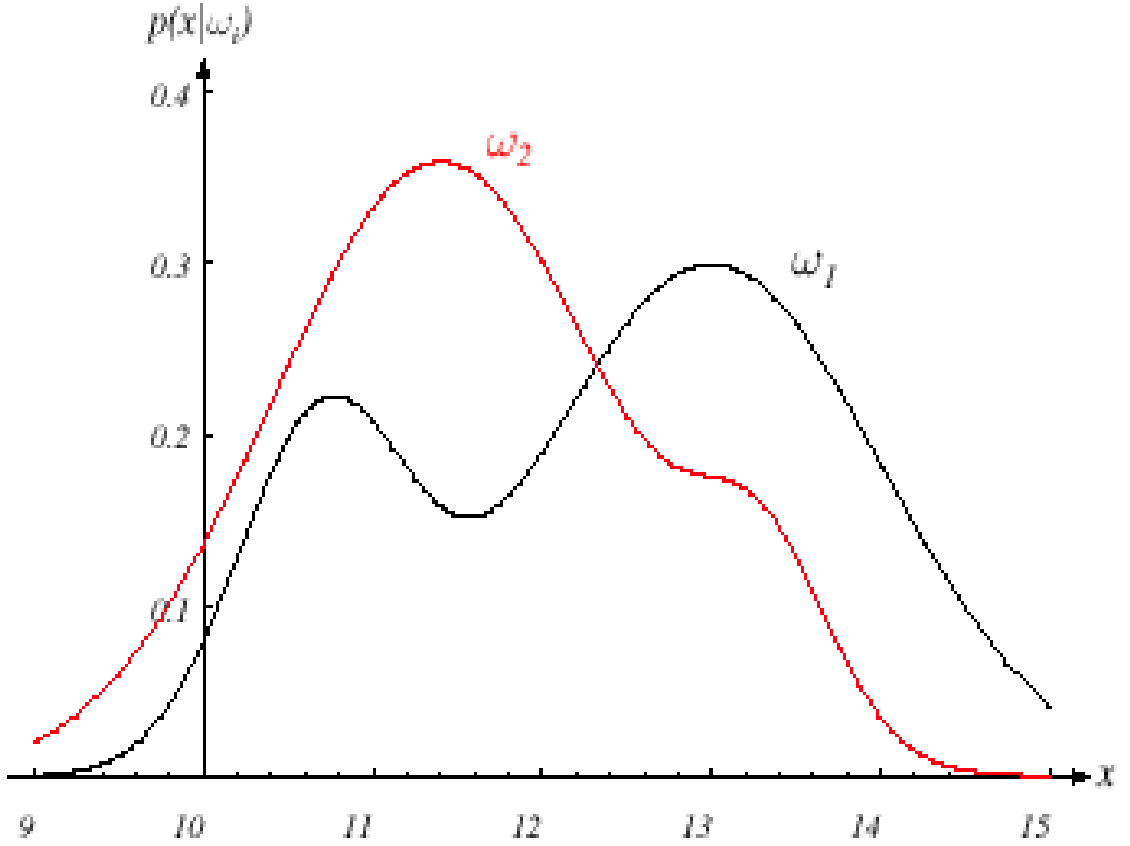

FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Regla de Decisión con variable X

Observamos ahora una característica X y tiene el valor x. Como influye esto en la determinación del estado del pez?

- X es una variable aleatoria que toma un valor x
- $p(x \mid \omega_1)$ y $p(x \mid \omega_2)$ probabilidades condicionadas a los estados
- $P(\omega_1)$ y $P(\omega_2)$ son las probabilidades a priori
- La probabilidad conjunta de x y ω *es*

$$p(\omega_j, x) = P(\omega_j | x).p(x) = p(x|\omega_j).P(\omega_j)$$

Regla de Decisión con variable X

Probabilidad del estado ω dado que se observa el valor x es llamada probabilidad a posteriori

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j) P(\omega_j)}{p(x)}$$

- Posterior = (Verosimilitud condicional x Prior) / Evidencia
- La evidencia es la densidad de la variable X

$$p_X(x) = \sum_{j=1}^{j=2} p(x \mid \omega_j) P(\omega_j)$$

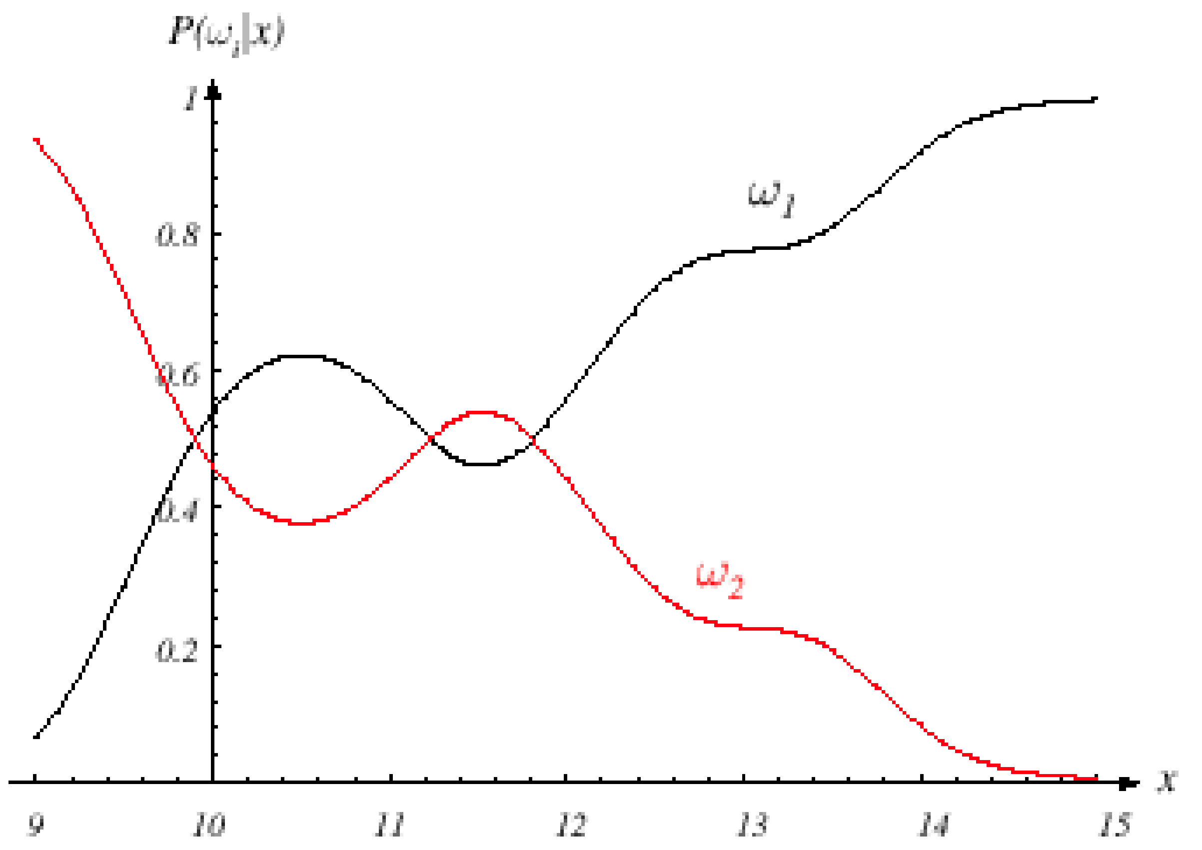

FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*.

Decisión dadas las probabilidades a posteriori

x es una observación:

si $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ ⇒ Decido estado natural = ω_1
si $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ ⇒ Decido estado natural = ω_2

Error

- Verdadero = ω_1 *Si me equivoco, decido por* ω_2 *es porque* $P(\omega_1 \mid x) < P(\omega_2 \mid x)$. Por lo cual $P(error \mid x) = P(\omega_1 \mid x)$
- Verdadero = ω_2 *Si me equivoco, decido por* ω_1 *es porque* $P(\omega_1 \mid x) > P(\omega_2 \mid x)$. Por lo cual $P(error \mid x) = P(\omega_2 \mid x)$
- *En ambos casos, la probabilidad a posteriori es la mas chica*

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

Regla de Bayes

- La regla de Bayes es la regla que produce el mínimo error promedio

Produce la regla basada en las probabilidades a posteriori el mínimo error promedio??

Es la regla basada en las probabilidades a posteriori la regla de Bayes??

Error promedio

- La probabilidad de error promedio es

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

- Y si *P(error | x)* es lo mas chico posible, la integral es la mas chica posible.

Error promedio

- Ahora, para cualquier regla, en cuanto observamos un x particular, $P(error \mid x)$ es :

  - $P(error \mid x) = P(\omega_1 \mid x)$ *si decidimos por* ω_2 *si era* ω_1
  - $P(error \mid x) = P(\omega_2 \mid x)$ *si decidimos por* ω_1 *si era* ω_2

- Por lo cual la regla cuyo error es

$$P(error \mid x) = min\,[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$

Es la regla con mínimo error promedio

Regla de Bayes

- En este contexto (conocemos probabilidades a priori y a posteriori)

La regla de Bayes es la regla basada en las probabilidades a posteriori

Regla de Bayes

- La regla de Bayes

si $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ Decido estado natural = ω_1
si $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ Decido estado natural = ω_2

- puede expresarse como

Decidir ω_1 *si* $p(x \mid \omega_1)\, P(\omega_1) > p(x \mid \omega_2)\, P(\omega_2)$;
en otro caso decidir ω_2

Generalización de las ideas precedentes

- Usar más de una característica
- Usar más de dos estados verdaderos
- Permitir acciones y no solo decidir el estado verdadero
- Introducir una función de pérdida que es mas general que la probabilidad de error

Generalizaciones

- Permitir otras acciones además que la clasificación
  - Permitir la posibilidad de rechazo:
  - Esto es, no tomar una decisión en casos cercanos o en malos casos!
- Introducir una función de pérdida :
  - La función de perdida dice cuanto cuesta cada acción que se toma

Generalizaciones

- Sea $\{\omega_1, \omega_2, \ldots, \omega_c\}$ el conjunto de c estados verdaderos (o “categorías”)

- Sea $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ el conjunto de posibles acciones

- Sea $\lambda(\alpha_i \mid \omega_j)$ la pérdida incurrida por tomar la acción α_i cuando el estado verdadero es ω_j

- X es un vector aleatorio y $p(x|\omega_j)$ es la probabilidad condicional de X cuando ω_j es verdadero

Posterior, verosimilitud, evidencia

- Posterior = (Verosimilitud x Prior) / Evidencia

$$\boxed{P(\omega_j \mid x) = \frac{p(x \mid \omega_j)P(\omega_j)}{P(x)}}$$

- Evidencia

$$\boxed{P(x) = \sum_{j=1}^{c} p(x \mid \omega_j)P(\omega_j)}$$

Riesgo condicional

- Sea $\lambda(\alpha_i \mid \omega_j)$ la pérdida incurrida por tomar la acción α_i cuando el estado verdadero es ω_j ,

- La pérdida esperada dada la observación *x* es

$$\boxed{R(\alpha_i \mid x) = \sum_{j=1}^{c} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x) \qquad i = 1, \ldots, a}$$

- Y se llama *Riesgo condicional*

Regla de Bayes

Se busca una regla $\alpha(x)$ que minimice el Riesgo

$$\boxed{R = \int R(\alpha(x)\,|\,x)\,p(x)\,dx}$$

Minimizar R ⟷ Minimizar $R(\alpha_i \mid x)$ *para* $i = 1,\ldots, a$

Por lo cual se busca un conjunto de acciones α_i para las cuales $R(\alpha_i \mid x)$ es mínimo

Riesgo de Bayes es el error de la Regla de Bayes

- Ese conjunto de acciones *$\alpha^*(x)$ es la regla de decisión de Bayes*

- Para ese conjunto de acciones $\alpha^*(x)$, el valor de la integral

$$\boxed{R^* = \int R(\alpha^*(x) \mid x) p(x) dx}$$

se llama riesgo de Bayes   y es el mejor desempeño posible!

Clasificación de dos categorías

- Reglas de decisión

$$\alpha_1 : \textit{decide por } \omega_1$$

$$\alpha_2 : \textit{decide por } \omega_2$$

- Perdida incurrida al decidir por ω_i cuando el verdadero estado natural es ω_j *es*

$$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$$

- Riesgo condicional:

$$R(\alpha_1 \mid x) = \lambda_{11} P(\omega_1 \mid x) + \lambda_{12} P(\omega_2 \mid x)$$
$$R(\alpha_2 \mid x) = \lambda_{21} P(\omega_1 \mid x) + \lambda_{22} P(\omega_2 \mid x)$$

Regla de Bayes

La regla de Bayes minimiza el riesgo condicional

La regla es la siguiente :

si $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

se toma la acción α_1: "decide por ω_1"

Si no se toma la acción α_2: "decide por ω_2"

Pruebe que minimiza el riesgo!!!

Regla de Bayes

Reemplazando….

“ Decidir por ω_1 si:

$$\boxed{\left(\lambda_{21}-\lambda_{11}\right)p(x\,|\,\omega_1)P(\omega_1) > \left(\lambda_{12}-\lambda_{22}\right)p(x\,|\,\omega_2).P(\omega_2)}$$

decidir por ω_2 en otro caso”

Regla de Bayes

Si además suponemos que $\lambda_{21} > \lambda_{11}$

- *Si ocurre que*

$$\boxed{\frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)} = T^*}$$

tome acción α_1 (decida por ω_1)

- *En otro caso tome acción α_2 (decida por ω_2)*

Propiedad de decisión optima

"Si el cociente de verosimilitudes excede el valor de un umbral T^* que es independiente del patrón ingresado x, se pueden tomar decisiones optimas"

Densidad Normal

- Densidad analíticamente manejable
- Densidad continua
- Muchos procesos son asintóticamente Gaussianos
- Caracteres escritos a mano, sonidos del habla se modelan como prototipos corruptos por un proceso aleatorio

$$\boxed{P(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],}$$

μ = esperanza de X

σ^2 = varianza

Densidad Normal

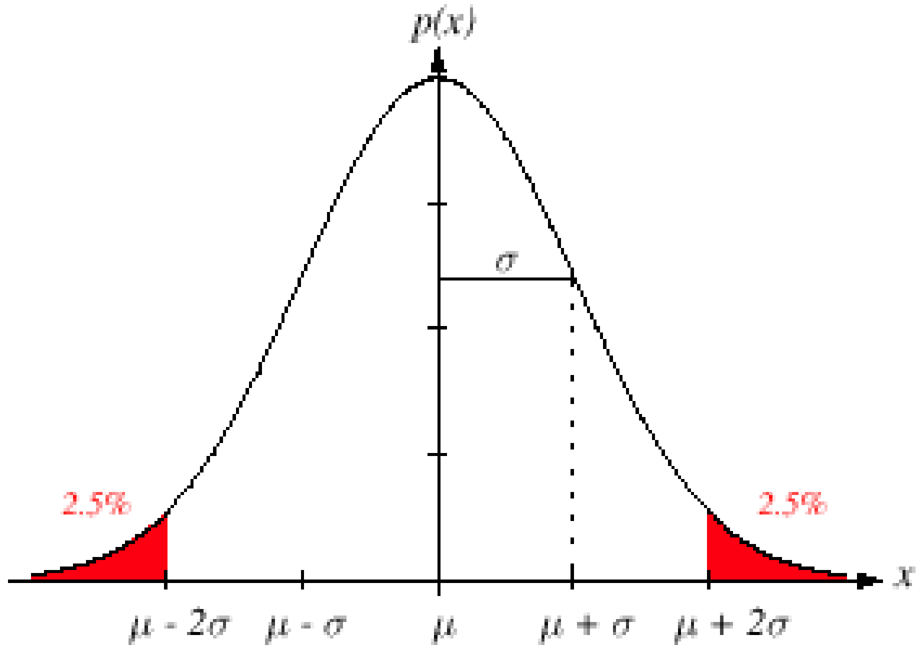

FIGURE 2.7. A univariate normal distribution has roughly 95% of its area in the range $|x - \mu| \leq 2\sigma$, as shown. The peak of the distribution has value $p(\mu) = 1/\sqrt{2\pi}\sigma$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

Ejemplo

Seleccionar la decisión optima cuando :

□ $= \{\omega_1, \omega_2\}$

$p(x \mid \omega_1)$ → N(2, 0.5) (distribución Normal)

$p(x \mid \omega_2)$ → N(1.5, 0.2)

$P(\omega_1) = 2/3$

$P(\omega_2) = 1/3$

$$\boxed{\lambda = \begin{bmatrix} 1 & 2 \\ 3 & 0.4 \end{bmatrix}}$$

La condición que deben cumplir los λ son $\lambda_{21} > \lambda_{11}$, $\lambda_{12} > \lambda_{22}$

Ejemplo

$$\frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

$$\frac{\dfrac{1}{\sqrt{2\pi}\,0.5}\exp\left[-\dfrac{1}{2}\left(\dfrac{x-2}{0.5}\right)^2\right]}{\dfrac{1}{\sqrt{2\pi}\,0.2}\exp\left[-\dfrac{1}{2}\left(\dfrac{x-1.5}{0.2}\right)^2\right]} > \frac{2-0.4}{3-1} \cdot \frac{2/3}{1/3}$$

Ejemplo

- *Si ocurre que*

$$2*\ln(0.2/0.5)-\left(\frac{x-2}{0.5}\right)^2+\left(\frac{x-1.5}{0.2}\right)^2>1.6$$

tome acción α_1 *(decida por* ω_1*)*

- *En otro caso tome acción* α_2 *(decida por* ω_2*)*