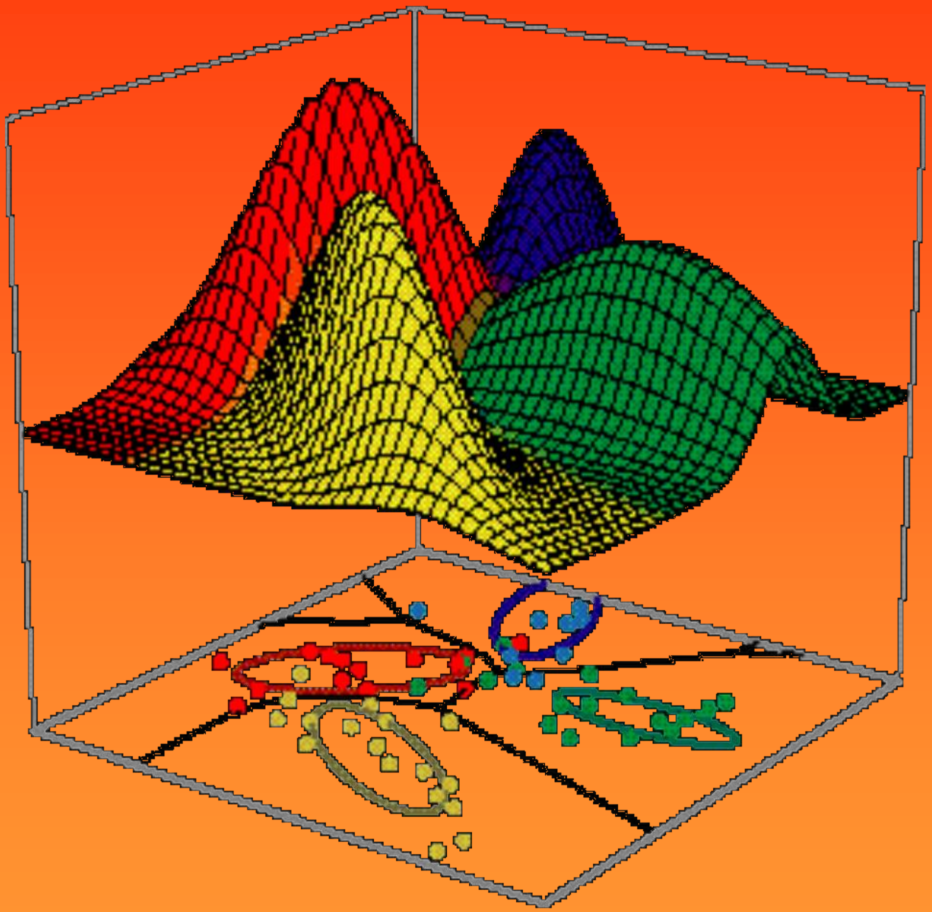


Pattern Classification

All materials in these slides were taken from Pattern Classification (2nd ed) by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000 **with the permission of the authors and the publisher**

Capitulo 2
Bayesian Decision Theory
(Sections 2.1-2.2)

Bayesian Decision Theory–Continuous Features

Introduccion

- Estado verdadero ω es una variable aleatoria

  - Dos estados verdaderos posibles: Salmon , Mero

- Probabilidades a priori
  - Pesca del Salmon $P(\omega_1)$
  - Pesca del Mero $P(\omega_2)$

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

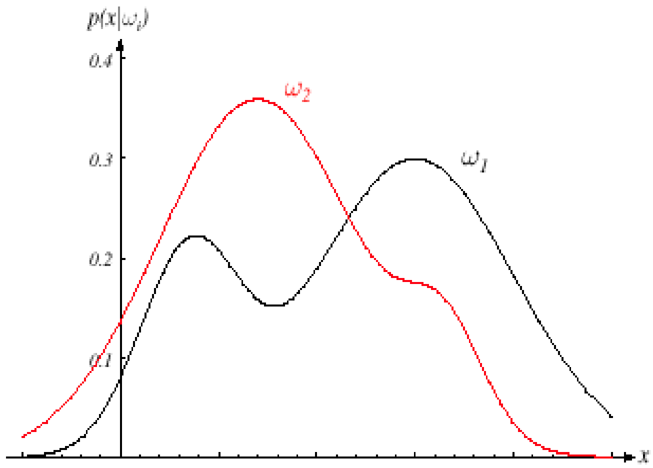
- Ejemplo: $P(\omega_1) = P(\omega_2)$ *equiprobables*

Regla de Decision solo con informacion a priori

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

Uso de la clase –informacion condicional probabilistica

- $p(x \mid \omega_1)$ and $p(x \mid \omega_2)$ describe la diferencia en tono de las escamas entre poblaciones de Mero y salmon



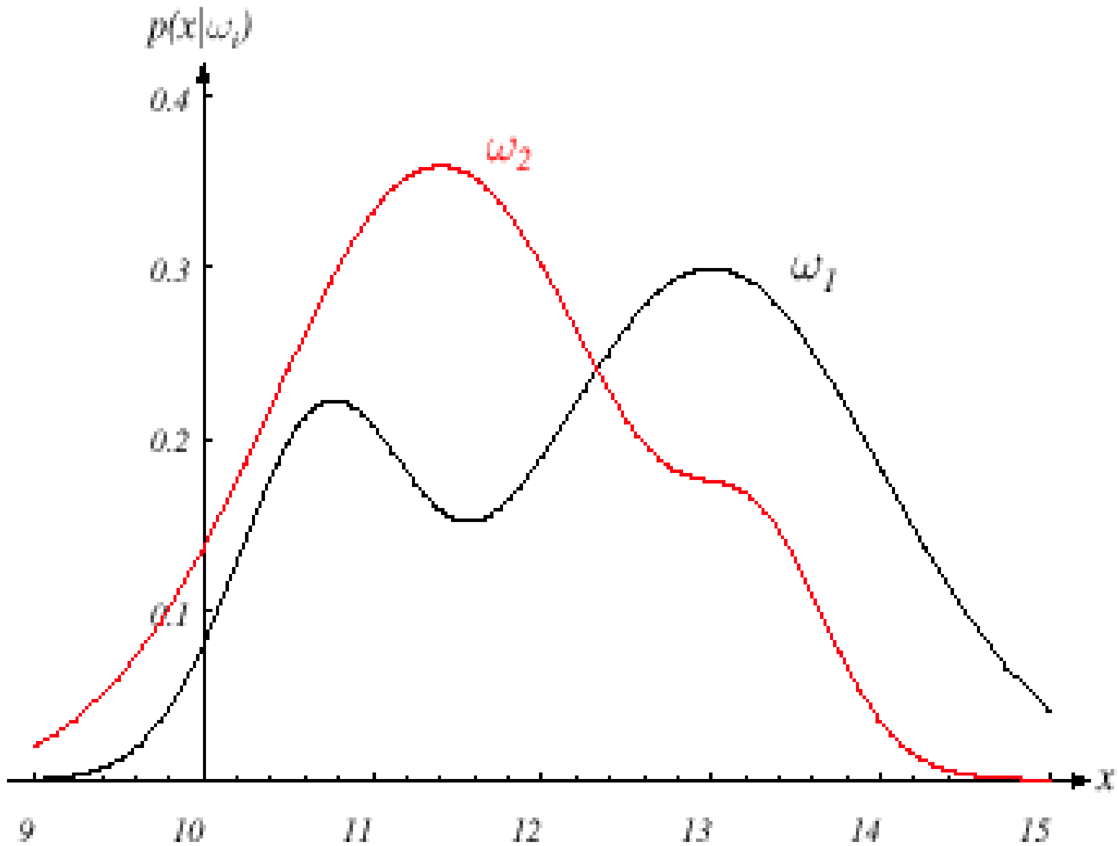


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two 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.

Posteriori, verosimilitud, evidencia

- P(x, ω_j) probabilidad conjunta de observacion y estado natural

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

- Posterior = (Verosimilitud x Prior) / Evidencia

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

  - En el caso de las dos categorias, la evidencia es

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

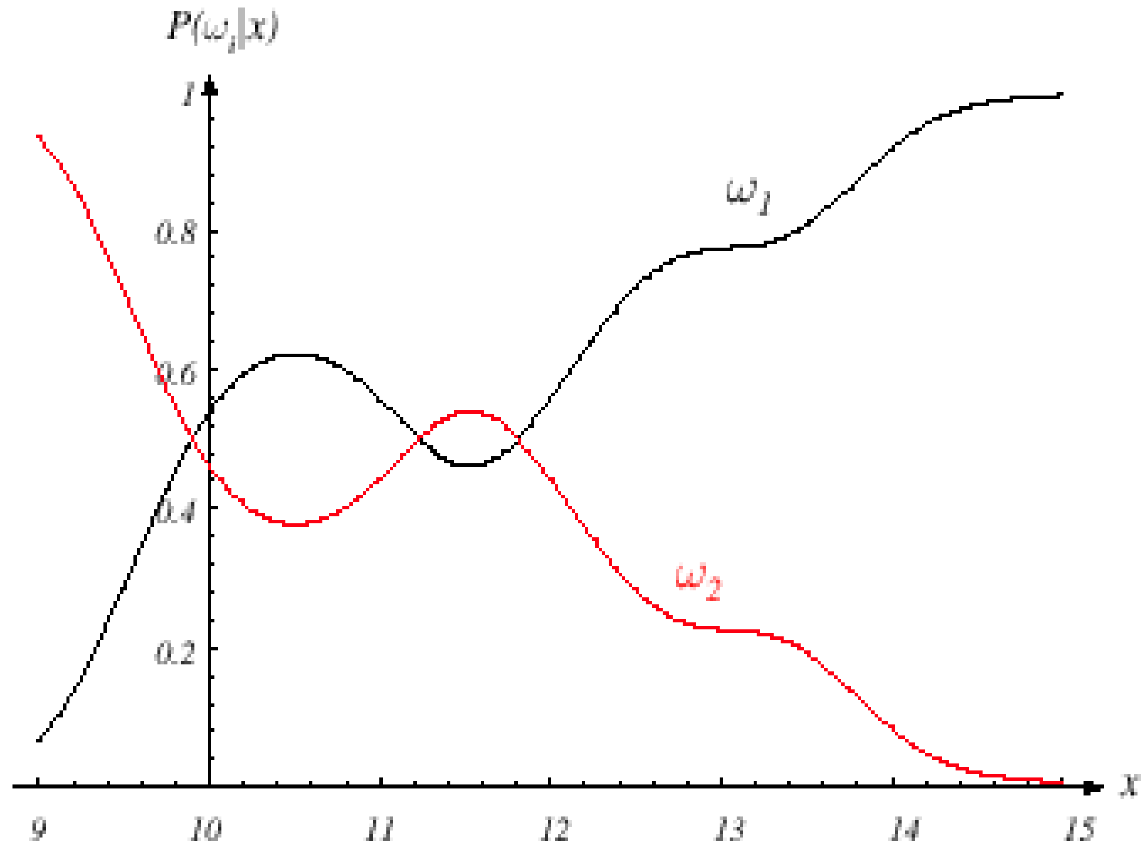


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly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

Regla de dadas las probabilidades a posteriori

x es una observacion:

si $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ → Verdadero estado natural = ω_1

si $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ → Verdadero Estado Natural = ω_2

Error

en cuanto observamos un x particular, la probabilidad de error es :

$$P(error \mid x) = P(\omega_1 \mid x) \text{ si decidimos por } \omega_2$$

$$P(error \mid x) = P(\omega_2 \mid x) \text{ si decidimos por } \omega_1$$

Regla de Bayes

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

(regla basada en las probabilidades a posteriori!!)

- Con esta regla

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

Produce esta regla el minimo error??

Produce esta regla el minimo error??

- La probabilidad de error promedio es

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

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

Con la regla de Bayes

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

por lo cual se garantiza que P(error) es minimo

- Descartando la evidencia p(x), la regla de Bayes puede expresarse como

Decidir por ω_1 *si*

$$p(x \mid \omega_1)\, P(\omega_1) > p(x \mid \omega_2) P(\omega_2);$$

en otro caso decidir por ω_2

- Si se conoce informacion adicional,
  - $p(x \mid \omega_1) = p(x \mid \omega_2)$, *entonces para ese x particular, entonces la decision depende enteramente de las probabilidades a priori .*
  - $P(\omega_1) = P(\omega_2)$ *entonces la decision se basa enteramente en las* $p(x \mid \omega_2)$

Teoria de Decision Bayesiana
Caracteristicas continuas

- Generalizacion de las ideas precedentes
  - Usar mas de una caracteristica
  - Usar mas de dos estados verdaderos
  - Permitir acciones y no solo decidir el estado verdadero
  - Introducir una funcion de perdida que es mas general que la probabilidad de error

- Permitir otras acciones ademas que la clasificacion permite la posibilidad de rechazo
- Se niega a tomar una decision en casos cercanos o en malos casos!
- La funcion de perdida dice cuanto cuesta cada accion que se toma

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

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

- Sea $\lambda(\alpha_l \mid \omega_j)$ la perdida incurrida por tomar la accion α_i cuando el estado verdadero es ω_j.

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

Posterior, verosimilitud, evidencia

- *Evidencia Conjunta*

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

- Posterior = (Verosimilitud x Prior) / Evidencia

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

Perdida esperada al tomar accion α_i

- Si $\lambda(\alpha_i \mid \omega_j)$ es la perdida incurrida por tomar la accion α_i cuando el estado verdadero es ω_j , la perdida esperada es

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

- La perdida esperada se llama usualmente Riesgo, y $R(\alpha_i \mid x)$ *es el Riesgo condicional*

- Se busca una regla que minimize el Riesgo completo

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

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

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

Seleccionar la accion α_i para la cual $R(\alpha_i \mid x)$ es minimo

→ R es minimo

R en ese caso se llama el riesgo de Bayes

el mejor desempeño posible!

- Clasificacion de dos categorias

α_1 : decide por ω_1

α_2 : decide por ω_2

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

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

Riesgo condicional:

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

La regla de Bayes es la siguiente :

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

se toma la accion α_1: “decide por ω_1”

Regla1

“ Decidir por ω_1 si:

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

decidir por ω_2 en otro caso”

Regla 2:

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

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

Entonces tome accion α_1 (decida por ω_1)

En otro caso tome accion α_2 (decida por ω_2)

Propiedad de decision optima

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

Densidad Normal

- Univariada
  - Densidad analiticamente manejable
  - Densidad continua
  - Muchos procesos son asintoticamente gausianos
  - Caracteres escritos a mano, sonidos del habla se modelan como prototipos corruptos por un proceso aleatorio

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

Donde:

μ = esperanza de x

σ^2 = varianza

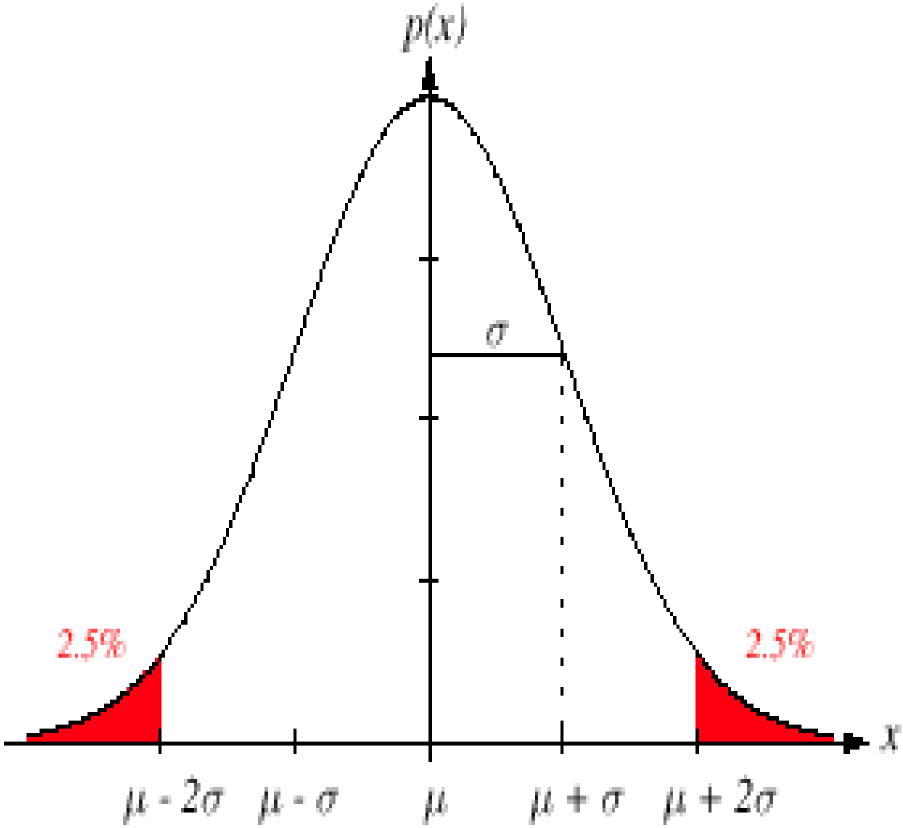


FIGURE 2.7. A univariate normal distribution has roughly 95% of its area in the range $|x - \mu| \leq 2\sigma$, as shown. The peak of the distribution has value $p(\mu) = 1/\sqrt{2\pi}\sigma$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John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

Ejemplo

Seleccionar la decision optima cuando :

$\Omega = \{\omega_1, \ \omega_2\}$

$p(x \mid \omega_1)$ → N(2, 0.5) (distribucion 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 condicion que deben cumplir los λ son $\lambda_{21} > \lambda_{11}$, $\lambda_{12} > \lambda_{22}$

Ejemplo

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

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

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