Universidade Federal do Paraná
Departamento de Informática

Reconhecimento de Padrões

Teoria da Decisão Bayesiana

Luiz Eduardo S. Oliveira, Ph.D.
http://lesoliveira.net

Teoria da Decisão Bayesiana

- Abordagem estatística fundamental em problemas de classificação.
- Quantificar o custo/benefício entre diferentes decisões de classificação usando probabilidades e custos associados a classificação.
  - Cada ação tem um custo associado.
  - O risco mais simples é o erro de classificação
  - Construir classificadores que minimizem o risco.

Terminologia

- Classes ω (variável aleatória)
  - ω_1 para robalo, ω_2 para salmão.
- Probabilidades a priori $P(\omega_1)$ and $P(\omega_2)$
  - Conhecimento a priori de pescar robalo ou salmão.
- Função de densidade probabilidade p(x)
  - Freqüência com a qual encontramos uma determinada característica
  - Evidências.

Terminologia

- Densidade de probabilidade condicional
  - $p(x/\omega_j)$ (Likelihood)
  - Frequencia com que encontramos uma determinada característica dado que a mesma pertence a classe ω_j

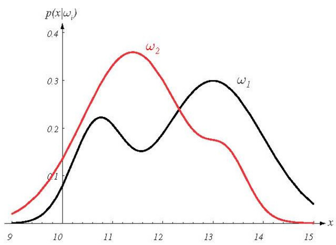


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two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

Terminologia

- Probabilidade a posteriori $P(\omega_j / x)$
  - Probabilidade que o peixe pertença a classe ω_j dado a característica x.
- Regra de decisão usando somente *priors*
  - ω_1 Se $P(\omega_1) > P(\omega_2)$; Senão ω_2 .
  - Essa regra nos faria tomar a mesma decisão todas as vezes.

Regra de Decisão usando Bayes

$$P(\omega_j / x) = \frac{p(x/\omega_j)P(\omega_j)}{p(x)} = \frac{likelihood \times prior}{evidence}$$

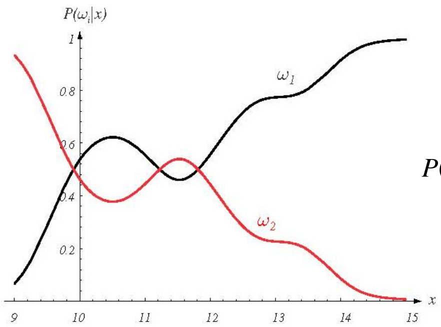
onde $p(x) = \sum_{j=1}^{2} p(x/\omega_j)P(\omega_j)$

Escolha ω_1 Se $P(\omega_1 |x) > P(\omega_2|x)$; senão escolha $\omega 2$

or

Escolha ω_1 if $p(x|\omega_1)P(\omega_1)>p(x/\omega_2)P(\omega_2)$ otherwise **decide** ω_2

Regra de Decisão usando Bayes



$$P(\omega_1) = \frac{2}{3} \qquad P(\omega_2) = \frac{1}{3}$$

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. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification.*

Probabilidade de Erro

- A probabilidade de erro usando Bayes é dada por
  - *P(error/x) = min[P(ω_1|x), P(ω_2|x)]*

Obtendo as probabilidades

- Essa teoria funciona somente quando conhecemos as funções.
- Abordagem objetiva
  - As probabilidades são obtidas através de experimentos
- Abordagem subjetiva
  - As probabilidades refletem um grau de confiança baseada em opinião ou conhecimento prévio.

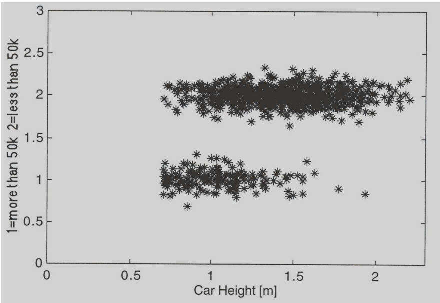
Exemplo

- Faça uma pesquisa no campus da PUCPR, perguntando valor e altura do carro que a pessoa possui
  - C1: preço > 20k
  - C2: preço < 20k
  - Característica X: Altura do carro.
- Usando Bayes, podemos calcular a probabilidade a posteriori.

$$P(C_i / x) = \frac{p(x / C_i) P(C_i)}{p(x)}$$

Exemplo (cont)

- Determinando *priors*
- Para cada carro, perguntar o preço e altura.
  - Por exemplo, 1209 carros
    - C1 = 221 e C2 = 988

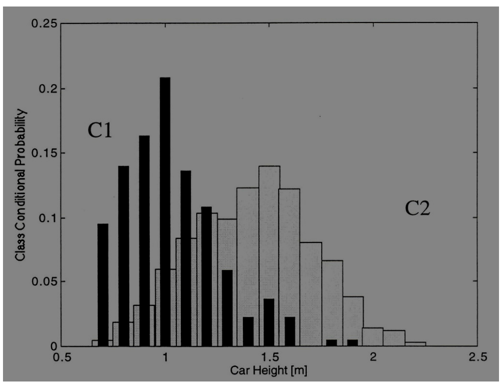


$$P(C_1) = \frac{221}{1209} = 0.183$$

$$P(C_2) = \frac{988}{1209} = 0.817$$

Exemplo (cont)

- Determinar a distribuição de probabilidade



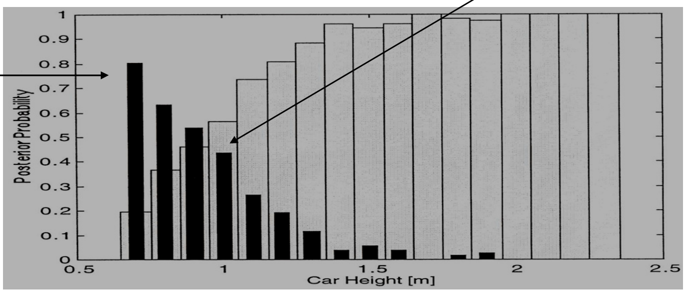
Exemplo (cont)

- Para cada barra do histrograma discretizado, calcular a probabilidade a posteriori.

$$P(C_1 / x = 1.0) = \frac{p(x = 1.0 / C_1)\, P(C_1)}{p(x = 1.0 / C_1)\, P(C_1) + p(x = 1.0 / C_2)\, P(C_2)} =$$

$$= \frac{0.2081 * 0.183}{0.2081 * 0.183 + 0.0597 * 0.817} = 0.438$$

Note qua deve Somar 1



Teoria Generalizada

- Uso de mais de uma característica
- Mais de duas classes
- Possibilita outras ações além da classificação (rejeição)
- Introduz uma função de erro mais genérica (*loss* function)
  – Associa custos com cada ação.

Terminologia

- Características formam um vetor $\mathbf{x} \in R^d$
- Conjunto finito de classes $\omega_1, \omega_2, \ldots, \omega_c$
- Conjunto finito de ações $\alpha_1, \alpha_2, \ldots, \alpha_l$
- *Loss function* $\lambda(\alpha_i / \omega_j)$
- A perda por tomar uma ação α_i quando a classificação for ω_j
- Bayes $$P(\omega_j / \mathbf{x}) = \frac{p(\mathbf{x} / \omega_j) P(\omega_j)}{p(\mathbf{x})}$$

onde $p(\mathbf{x}) = \sum_{j=1}^{c} p(\mathbf{x} / \omega_j) P(\omega_j)$ $(\textit{scale factor})$

Minimização do Risco

- Risco condicional (*Expected loss*) de tomar uma ação α_i

$$R(a_i / \mathbf{x}) = \sum_{j=1}^{c} \lambda(a_i / \omega_j) P(\omega_j / \mathbf{x})$$

- Esse risco pode ser minimizado selecionando-se uma ação que minimiza o risco condicional.

Minimização do Risco

- A regra de Bayes que minimiza R
  - Computando $R(\alpha_i /\mathbf{x})$ para cada α_i dado um **x.**
  - Escolher uma ação α_i com o mínimo $R(\alpha_i /\mathbf{x})$

Exemplo
Problema com duas Classes

- Duas possíveis ações
  - α_1 corresponde a decidir por ω_1
  - α_2 corresponde a decidir por ω_2
- Notação

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

- Os riscos condicionais são

$$R(a_1/\mathbf{x}) = \lambda_{11}P(\omega_1/\mathbf{x}) + \lambda_{12}P(\omega_2/\mathbf{x})$$
$$R(a_2/\mathbf{x}) = \lambda_{21}P(\omega_1/\mathbf{x}) + \lambda_{22}P(\omega_2/\mathbf{x})$$

onde λ_{ij} is equivalent to $\lambda(a_i/\omega_j)$

Exemplo
Problema com duas Classes

- Regra de decisão
  - ω_1 Se $R(\alpha_1 / \mathbf{x}) < R(\alpha_2 / \mathbf{x})$;
  - ω_2 Caso contrário