Resumo de Aprendizagem de Máquina 2014-2

Eduardo M. B. de A. Tenório
embat@cin.ufpe.br

CIn-UFPE

Resumo

Este documento tem por finalidade ser um resumo dos assuntos abordados na disciplina Aprendizagem de Máquina *do período 2014-2 do CIn-UFPE, ministrada pelos professores Francisco Carvalho e Teresa Ludermir. A maioria do documento referencia o livro "Pattern Classification", de Duda, Hart e Stork. Os códigos utilizados como exercício de fixação encontram-se em* https://github.com/embatbr/aprendizagem.

1 Teoria da Decisão Bayesiana

Teoria da Decisão Bayesiana é uma abordagem estatística para a classificação de padrões, baseada em quantificar os tradeoffs associados a tomar uma determinada decisão (classificar) utilizando probabilidade e considerando os custos associados.

O **estado natural** é denotado por ω, de modo que $\omega = \omega_i$, para $i = 1, 2, ..., c$, significa que o exemplo foi classificado como pertencente à classe ω_i. Cada uma dessas classes possui uma **probabilidade a priori** $P(\omega_i)$, com

$$\sum_{i=1}^{c} P(\omega_i) = 1, \qquad (1)$$

refletindo o conhecimento prévio da chance de um elemento da classe ω_i aparecer. A **regra de decisão** seria: Decida ω_i para $\max_i P(\omega_i)$. Neste caso a classe ω_i sempre é escolhida e a probabilidade de erro é dada por:

$$P_{err}(\omega_i) = 1 - P(\omega_i). \qquad (2)$$

Utilizando uma característica x que seja contínua e aleatória, sua **densidade de probabilidade estado-condicional** é dada por $p(x|\omega)$. Logo, a diferença entre $p(x|\omega_i)$ e $p(x|\omega_j)$ descreve a diferença da característica x entre as populações das classes ω_i e ω_j.

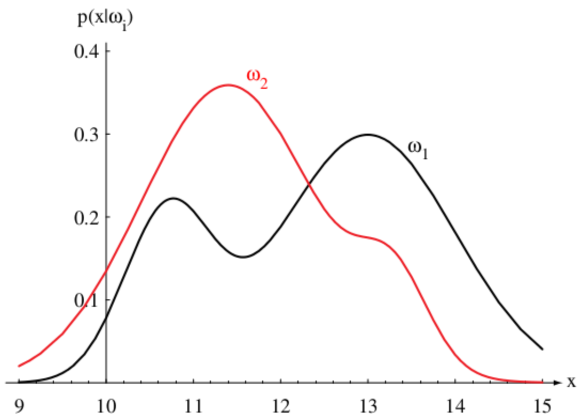

Figura 1: Para $\omega_i = \omega_2$, é mais frequente observar x entre 11 e 12 que $x = 13$ (valor mais provável se $\omega_i = \omega_1$).

Sabendo $P(\omega_i)$ e $p(x|\omega_i)$, e medindo um valor x, a probabilidade conjunta de achar um padrão na classe ω_i e com x é dado por: $p(\omega_i, x) = P(\omega_i|x)p(x) = p(x|\omega_i)P(\omega_i)$, que pela **fórmula de Bayes** fica:

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}, \tag{3}$$

com a evidência para c classes

$$p(x) = \sum_{j=1}^{c} p(x|\omega_j)P(\omega_j). \tag{4}$$

A fórmula de Bayes pode ser expressa em português como

$$posteriori = \frac{verossimilhanca \times priori}{evidencia}. \tag{5}$$

Fig. (2) mostra a probabilidade a posteriori das classes ω_1 e ω_2 para um conjunto de valores de x. A regra de decisão muda para: Decida ω_i se ω_i minimiza $P(erro|x)$, onde

$$P(erro|x) = \sum_{j \neq i} P(\omega_j|x), \tag{6}$$

ou simplesmente $P(erro|x) = 1 - P(\omega_i|x)$. Então a regra torna-se: Decida ω_i para $\max_i P(\omega_i|x)$

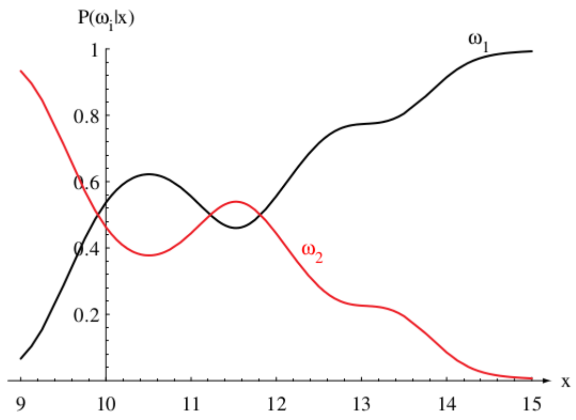

Figura 2: Probabilidades a posteriori para $P(\omega_1) = \frac{2}{3}$ e $P(\omega_2) = \frac{1}{3}$, e para as densidades de probabilidade estado-condicional mostradas em Fig. (1).

Esta regra minimiza a probabilidade média de erro, dada por

$$P(erro) = \int_{-\infty}^{\infty} P(erro|x)p(x)dx. \tag{7}$$