



Análise e classificação: sinais biológicos

Prof. Sérgio S Furuie

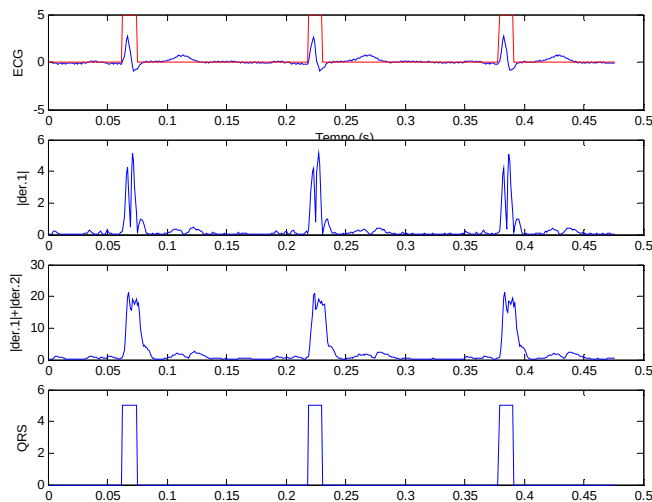


Mobilização/motivação

- Subsídios objetivos para a tomada de decisão
- Auxílio para a interpretação médica
- Exemplos:
 - Área de QRS
 - Variabilidade de IRR
 - Máxima taxa de variação
 - Classe



Deteção do QRS



$y1=|1a. \text{difer.}|$
 $y2=|smooth \ y1|$
 $y3=|der. \ y2|$
 $y4=y2+y3$
 limiar em $y4$



Extração de características

- Atributos relevantes e discriminativos
- Exemplos:
 - Área do QRS
 - Duração do QRS
 - Intervalo RR (IRR)
 - Correlação com uma referência
 - ...



Conceitos importantes

Correlação [sinais $x(t)$ e $y(t)$]

- Correlação: $\text{corr}(x,y)$
- Função correlação cruzada: $\text{corr}_{xy}(\tau)$
- F. autocorrelação: $\text{corr}_{xx}(\tau)$

$$\text{corr}(x, y) = \int_{-\infty}^{\infty} x(t) \cdot y(t) \cdot dt$$

$$\cong \sum_0^{N-1} x(n) \cdot y(n)$$

$$\text{corr}_{xy}(\tau) = \int_{-\infty}^{\infty} x(t) \cdot y(t + \tau) \cdot dt$$

$$\text{corr}_{xx}(\tau) = \int_{-\infty}^{\infty} x(t) \cdot x(t + \tau) \cdot dt$$

Coefficiente de correlação

- Normalizado entre [0,1]
- Extraídas as médias
- Normalizado pelo d. padrão

$$\rho_{xy} = \frac{\int_{-\infty}^{\infty} (x(t) - \mu_x) \cdot (y(t) - \mu_y) \cdot dt}{\sigma_x \cdot \sigma_y}$$

$$\cong \frac{\sum_0^{N-1} (x(n) - \mu_x) \cdot (y(n) - \mu_y)}{\sigma_x \cdot \sigma_y}$$



Classificação não-supervisionada

- Clusters
 - K-means
 - Fuzzy c-means
 - SOM (Kohonen)



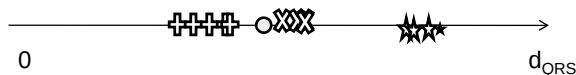
Classificação supervisionada

- Existem amostras de cada classe (C) tornando possível caracterizar parcialmente cada classe
 - Classificação direta
 - NN-Nearest-neighbor
 - K-NN: classificação pela maioria
 - Definição de distâncias:
 - Euclidiana
 - estatística (Mahalanobis)
 - Treinamento de algoritmos
 - Classificadores otimizados
 - Linear de Fischer
 - Bayesiano
 - SVM
 - Redes neurais (backpropagation)
 - Avaliação de algoritmos

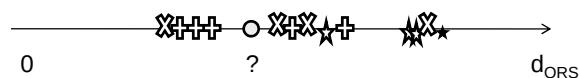


Classificação de padrões: 1 atributo

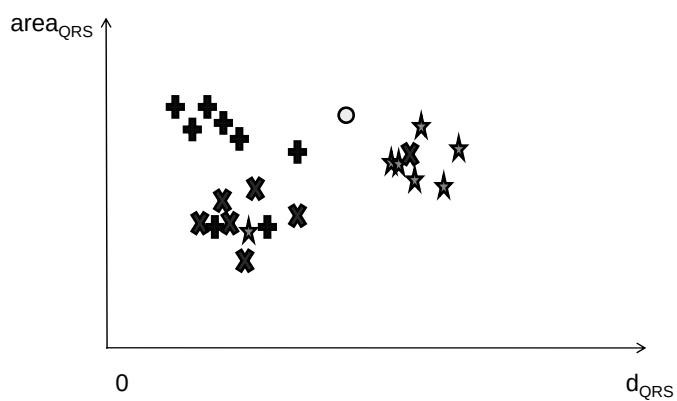
Ideal, desejada



Realidade:

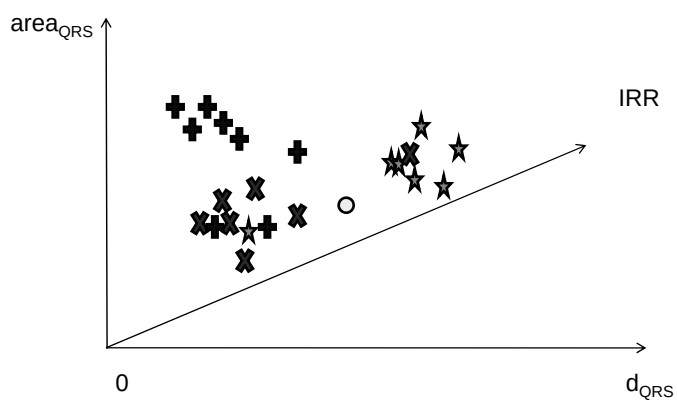


Classificação de padrões: 2 atributos



9

P atributos, C classes



Clusters (nuvens)

10



Classificação NN-Nearest Neighbor

1. Seja $S=\{s_i, i=1,N\}$ as amostras e $c(s)$ a classe do elemento s , $c:S \rightarrow \{1,2, ..C\}$
2. Determinar a distância do vetor x a todos os elementos em S
3. $c(x) \leq c(s_{NN})$

distância euclideana

$$\vec{s}_i = [s_{i1} \quad s_{i2} \quad \dots \quad s_{ip}]^T$$

$$\vec{x} = [x_1 \quad x_2 \quad \dots \quad x_p]^T$$

$$d^2(\vec{s}_i, \vec{x}) = \sum_{k=1, P} (s_{ik} - x_k)^2$$

distância estatística

$$d^2(\vec{s}_i, \vec{x}) = (\vec{s}_i - \vec{x})^T \cdot \text{Cov}^{-1} \cdot (\vec{s}_i - \vec{x})$$

Cov : matriz de covariância de interesse



Classificação k-NN

1. Definido: k
2. Obter os k vizinhos mais próximos
3. Classe de x será aquela com maioria entre os k vizinhos



Exemplo de aplicação: arritmias

- Tendo-se amostras dos diversos tipos de arritmias:
 - Supra-ventricular
 - Bigeminada
 - ...
- Obter conjunto de atributos para caracterizar os clusters de cada classe
- Aplicar classificador, p. ex. k-NN p/ decisão



Exemplo: arritmias em tempo real

- Usar trecho inicial, sob supervisão humana, para “aprender” padrão de QRS normal ($k=1$)
- Para cada novo QRS, comparar (correlação) com o normal e as demais classes existentes.
 - Caso difira substancialmente de todas, criar nova classe (arritmia tipo $k+1$)
 - Se pertencer a uma certa classe, atualizar o padrão de forma adaptativa.



Classificador Linear de Fischer

Dadas 2 classes A e B com alta dimensionalidade, como obter um classificador linear otimizado?

d: atributos

- Obter a combinação linear (pesos w) que maximiza a separabilidade entre as 2 classes
- Obter o ponto de decisão (threshold)

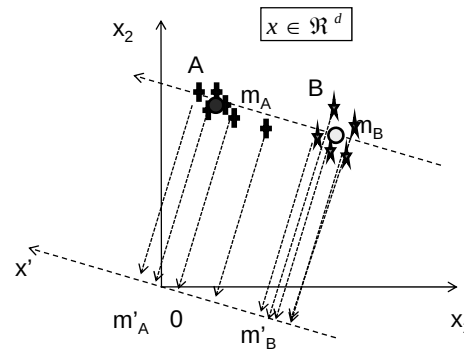
$$\sigma_{A',B'} = \frac{|m'_{A'} - m'_{B'}|^2}{S'_{A'} + S'_{B'}}$$

$$m'_{A'} = \frac{1}{N_{A'}} \sum_{x \in A'} x'$$

$$m'_{B'} = \frac{1}{N_{B'}} \sum_{x \in B'} x'$$

$$S'_{A'} = \sum_{x \in A'} (x' - m'_{A'})^2$$

$$S'_{B'} = \sum_{x \in B'} (x' - m'_{B'})^2$$



$$\sigma_{A',B'} = \frac{|m'_{A'} - m'_{B'}|^2}{S'_{A'} + S'_{B'}}$$

$$m'_{A'} = \frac{1}{N_{A'}} \sum_{x \in A'} x'$$

$$m'_{B'} = \frac{1}{N_{B'}} \sum_{x \in B'} x'$$

$$S'_{A'} = \sum_{x \in A'} (x' - m'_{A'})^2$$

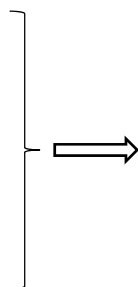
$$S'_{B'} = \sum_{x \in B'} (x' - m'_{B'})^2$$

Seja:

$$x \in \mathbb{R}^d$$

$$m_A = \frac{1}{N_A} \sum_{x \in A} x$$

$$m_B = \frac{1}{N_B} \sum_{x \in B} x$$



$$\sigma_{A,B}(w) = \frac{w^T S_E w}{w^T S_I w}$$

$$\max_w \sigma_{A,B}(w) \Rightarrow S_I w = m_A - m_B$$

$$w = S_I^{-1} (m_A - m_B)$$

S_E : espalhamento entre-classes
 S_I : espalhamento intra-classes

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

$$x' = w^T \cdot x$$

$$m'_A = \frac{1}{N_A} \sum_{x \in A} w^T \cdot x = w^T \cdot m_A$$

$$m'_B = w^T \cdot m_B$$

$$S'_A = \sum_{x \in A} (w^T \cdot x - w^T \cdot m_A)^2 = \sum_{x \in A} w^T \cdot (x - m_A) \cdot (x - m_A)^T \cdot w$$

$$= w^T \cdot \left[\sum_{x \in A} (x - m_A) \cdot (x - m_A)^T \right] \cdot w$$

$$S'_B = w^T \cdot \left[\sum_{x \in B} (x - m_B) \cdot (x - m_B)^T \right] \cdot w$$

$$|m'_A - m'_B|^2 = \left| w^T (m_A - m_B) \right|^2 =$$

$$w^T (m_A - m_B) \cdot (m_A - m_B) \cdot w =$$

$$= w^T \cdot S_E \cdot w$$

$$S'_A + S'_B = w^T \cdot \left[\sum_{x \in A} (x - m_A) \cdot (x - m_A)^T + \sum_{x \in B} (x - m_B) \cdot (x - m_B)^T \right] \cdot w$$

$$= w^T \cdot S_I \cdot w$$

$$S_E = (m_A - m_B) \cdot (m_A - m_B)^T$$

singular $\Rightarrow S_E \cdot w = K \cdot (m_A - m_B)$

$$S_I = \left[\sum_{x \in A} (x - m_A) \cdot (x - m_A)^T + \sum_{x \in B} (x - m_B) \cdot (x - m_B)^T \right]$$

$$\sigma_{A,B}(w) = \frac{w^T S_E w}{w^T S_I w}$$

EPUSP PTC/ LEB - S.Furuie 17

PTC 2456 – Proc. Sinais Biomédicos
Análise, quantificação, classificação de sinais biol.

Obtendo o ponto de decisão: ML

T: valor de y com igual possibilidade de ser classe A ou B

$$P(A | y) = P(B | y)$$

$$P(A | y) = \frac{p(y | A) \cdot P(A)}{p(y)}$$

$$P(B | y) = \frac{p(y | B) \cdot P(B)}{p(y)}$$

$$\therefore p(T | A) \cdot P(A) = p(T | B) \cdot P(B)$$

se Gaussianas com mesma variância ,

$$p(y | A) = \frac{1}{\sqrt{2\pi} \sigma_A} \exp\left(-\frac{(y - m'_A)^2}{2\sigma_A^2}\right)$$

$$P(A) \cdot \exp\left(-\frac{(T - m'_A)^2}{2\sigma^2}\right) = P(B) \cdot \exp\left(-\frac{(T - m'_B)^2}{2\sigma^2}\right)$$

$$T = \frac{m'_A + m'_B}{2} + \frac{\sigma^2}{m'_A - m'_B} \ln\left(\frac{P(B)}{P(A)}\right)$$

se $P(A) = P(B) \Rightarrow T = \frac{m'_A + m'_B}{2}$

EPUSP PTC/ LEB - S.Furuie 18



Avaliação

- Poucos casos classificados
 - Leave-one-out
 - Treinamento com N-1 casos
 - Avaliação com o elemento restante
- Amostras em grande número
 - Conjunto para treinamento
 - Conjunto independente para avaliação



Bibliografia

- Biomedical Signal Analysis. R.M. Rangayyan. Wiley Interscience, 2002