

Atributos globais

- Contornos
 - Descritores geométricos
 - Parâmetros de elementos geométricos
 - Retas → coeficiente angular
 - Segmentos → extremos
 - Círculos → centro, raio
 - Elipses → centro, semi-eixos
 - Polígonos → vértices
 - Curvas → curvatura

Descritores de regiões

- Área
- Perímetro
- Centróide (baricentro)
- Momentos
- Eixos principais
- Caixas envolventes (bounding box), fecho convexo
- Descritores topológicos
- Outros (círculo ou elipse de área equivalente, retângulo de área e orientação equivalente, etc)

Atributos invariantes

- A vantagem de se ter atributos invariantes a transformações de translação, escala e rotação é garantir maior robustez ao reconhecimento de padrões
 - Essas transformações são ditas geométricas
 - Seria desejável outros tipos de invariância, por exemplo, a transformações radiométricas (mudanças de escalas de intensidade, cromaticidade, transformações de histograma, filtragens, etc), todavia elas são ainda desconhecidas em grande parte.

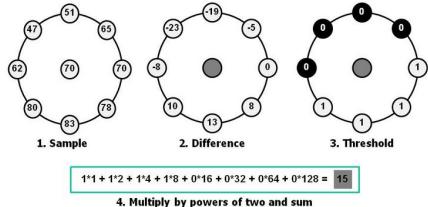
Local Binary Pattern

- Em 1990, L. Wang e D.C. He introduzem a ideia do expectro de textura em:
 - Texture classification using texture spectrum- Pattern Recognition, 23 (1990), pp. 905–910
- Em 1996, T. Ojala, M. Pietikäinen e D. Hardwood modificam a ideia para os LBP em:
 - A comparative study of texture measures with classification based on featured distributions Pattern Recognition, 29 (1996), pp 51–59

Local Binary Pattern

The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \quad s(x) = \begin{cases} 1, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$



Local Binary Pattern

Face image The face image is divided into blocks LBP histogram from each block Feature histogram

Classificação

- Pode-se fazer uma classificação após o cálculo dos histogramas de atributos

Aprendizado estatístico e computacional

Aprendizado não-supervisionado

- É um método popular de análise de dados
- O objetivo desta análise é encontrar estruturas geométricas nos dados
- A estrutura geométrica, se existir, vai pertencer ao espaço formado por um conjunto de medidas ou categorias que compõe os dados

Aprendizado não-supervisionado

- Atributo, ou característica, é o nome que se dá a um valor numérico (medida), ou categórico, associado a um elemento do conjunto de interesse
- Dentre os métodos de análise exploratória de dados, a análise de agrupamentos tornou-se o método mais famoso

Aprendizado não-supervisionado

- O nome **análise de agrupamentos, ou conglomerados**, vem do inglês “**cluster analysis**”.
- Depende da boa escolha do **número de conglomerados e da função de similaridade** entre os pontos no espaço de atributos

Análise de conglomerados

- Por exemplo:**
 - Numa certa análise populacional, podemos estar interessados na altura e comprimento de um animal

Análise de conglomerados

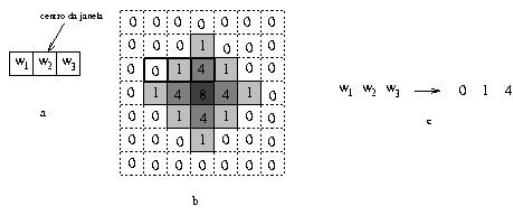
- Em certas análises, os dados podem estar categorizados. Por exemplo: alto(a), normal(a), baixo(a). Para cada categoria, associa-se normalmente um valor numérico.

Características sobre Imagens

- Pode ser, simplesmente, o valor de um ponto da imagem, ou o conjunto de valores vistos através de uma janela/máscara W
- O mais natural é tomar medidas sobre atributos locais, ou globais, da imagem, por exemplo: largura de um objeto, número de objetos em uma certa região, etc

Análise de conglomerados

- Exemplo usando o valor da imagem



Análise de conglomerados

- Tomando como características os valores vistos através de W, mais o número de vezes que eles apareceram, formamos o conjunto de dados de treinamento

W_1	W_2	W_3	N
0	0	0	14
0	0	1	3
0	1	0	4
1	0	0	4
0	1	4	3
4	1	0	3
1	4	1	2
1	4	8	1
8	4	1	1
4	8	4	1

Análise de aglomerados

- Como medimos o quanto similar são os pontos no espaço de características?
 - Resposta óbvia: defina uma métrica nesse espaço
 - Dois pontos no espaço de características que pertençam a mesma classe devem ser mais próximos que dois pontos que pertençam a classes diferentes

Distância versus similaridade

- Distância é uma função d que satisfaz:

$$d(a, b) \geq 0$$

$$d(a, b) = d(b, a)$$

$$d(a, a) = 0$$

$$d(a, b) = 0 \Leftrightarrow a = b$$

$$d(a, c) \geq d(a, b) + d(b, c)$$

Distância versus similaridade

- Similaridade é uma função d que satisfaz:

$$d(a, b) \geq 0$$

$$d(a, b) = d(b, a)$$

$$d(a, c) \geq d(a, b)$$
- Se a é mais similar a b que a c

Análise de aglomerados

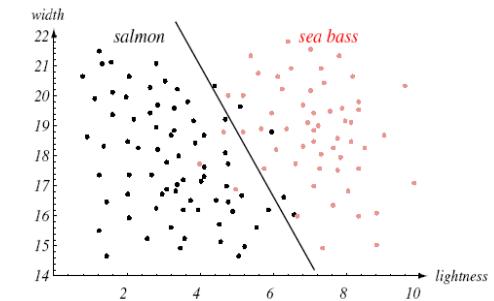
- Porém, qual a métrica adequada para medir similaridades de atributos com características diferentes?
- Por exemplo: altura X peso, quantidade de A, T, C, G X proteína, tRNA, mRNA

Análise de aglomerados

- Em PDI, é comum termos esse tipo de diferença de métricas



Análise de aglomerados

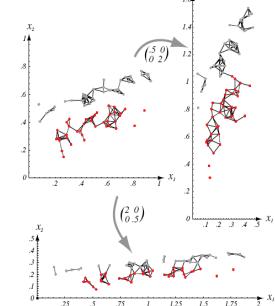


Análise de aglomerados

- Uma forma de comparar quantidades de dimensões diferentes é fazer uma transformação de escala
- Outra forma é normalizá-las

Transformações de escala

- Transformação de escala
 - Os aglomerados devem ser invariantes a esse tipo de transformação
 - Não deve-se fazer se a transformação não for natural, ou sem sentido



Transformações de escala

- Existem várias formas de fazer normalização de dados. A mais comum faz-se:
 - Subtrai-se a média da variável (atributos)
 - Escala-se a variável para que tenha variância unitária

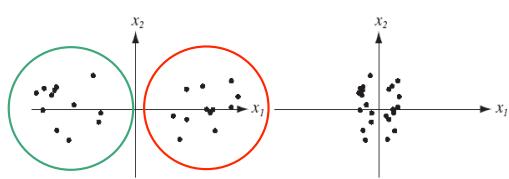
$$x' = \frac{x - \mu}{\sigma^2}$$

Medidas de similaridade

- A idéia desta transformação é que atributos cujos valores são altos deixam de dominar nos cálculos de distância
- Porém, deve-se fazer isso apenas se o espalhamento dos valores é devido a uma variação aleatória normal
- Se a variação é devida a presença de subclasse, pode-se prejudicar a aglomeração

Transformações de escala

- Exemplo de onde não é adequado usar a normalização de variáveis



Medidas de similaridade

- Ao invés de fazer transformações, pode-se usar medidas de distância diferentes
- Distância de Minkowski

$$d(x, x') = \left(\sum_{k=1}^d |x_k - x'_k|^q \right)^{\frac{1}{q}}$$

- $q = 1$, city block ; $q = 2$, euclidean

Medidas de similaridade

- A distância de Mahalanobis leva em consideração a covariância entre as variáveis para calcular a similaridade
- Assim, evitam-se os problemas de escala devido a distância euclideana
- Na métrica euclideana, o lugar geométrico dos pontos equidistantes de um certo ponto é uma esfera
- Nessa métrica, a esfera é esticada para corrigir as diferentes escalas e levar em conta a correlação entre as variáveis

Medidas de similaridade

- A distância de Mahalanobis pode ser escrita como:

$$d(x, x')^2 = (x - \mu)^t \Sigma^{-1} (x - \mu)$$

- Onde Σ^{-1} é a matriz de covariância das variáveis. Cada elemento de Σ^{-1} é dado por:

$$\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

Distância de Mahalanobis

- A matriz de covariância é simétrica e positivo semi-definida

$$x^* \Sigma x > 0$$

- Onde x^* é o transposto conjugado de x
- O que acontece com Σ quando as variáveis são independentes?

Distância de Canberra

- A distância de Canberra pode ser escrita como:

$$d(x, x') = \sum_{i=1}^d \frac{|x_i - x'_i|}{|x_i| + |x'_i|}$$

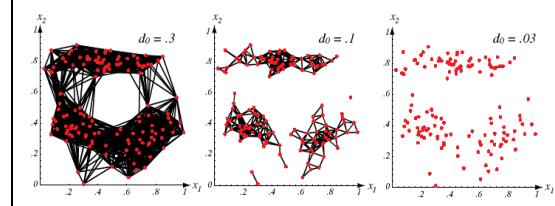
- Ela é sensível para valores próximos da origem

Análise de aglomerados

- Normalmente, diz-se que dois pontos pertencem ao mesmo cluster se sua distância é menor que um certo limiar
- A escolha desse valor é crítica
- Se o valor é alto demais, todos os pontos vão ser atribuídos ao mesmo cluster
- Se o valor é baixo demais, cada ponto será um cluster diferente

Análise de aglomerados

- O valor adequado é um pouco maior que a distância intra-cluster e um pouco menor que a distância entre-cluster típicas



Análise de aglomerados

- O algoritmo de cluster hierárquico é uma das implementações mais simples
- Além da função de similaridade, é preciso escolher como a distância entre um ponto e um cluster
- Algumas possibilidades: single-linkage, complete-linkage, ...

Algoritmo iterativo

- Dada uma função critério, o problema de aglomeração passa a ser um problema de otimização discreta
- Como o conjunto de exemplos é finito, o número de possíveis partições também é e o ótimo pode ser achado por busca exaustiva, em geral, num algoritmo iterativo
- Parte-se de um particionamento inicial e vai-se movendo exemplos de um cluster para o outro para tentar minimizar a função critério
- Essa abordagem, em geral, garante ótimos locais, mas não globais

Algoritmo iterativo

■ Algorithm 3. (Basic Iterative Minimum-Squared-Error Clustering)

```

1 begin initialize  $n, c, \mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_c$ 
2     do randomly select a sample  $\hat{\mathbf{x}}$ 
3          $i \leftarrow \arg \min_{l'} \|\mathbf{m}_{l'} - \hat{\mathbf{x}}\|$  (classify  $\hat{\mathbf{x}}$ )
4         if  $n_i \neq 1$  then compute
5              $\rho_j = \begin{cases} \frac{n_j}{n_j+1} \|\hat{\mathbf{x}} - \mathbf{m}_j\|^2 & j \neq i \\ \frac{n_j}{n_j-1} \|\hat{\mathbf{x}} - \mathbf{m}_i\|^2 & j = i \end{cases}$ 
6             if  $\rho_k \leq \rho_j$  for all  $j$  then transfer  $\hat{\mathbf{x}}$  to  $\mathcal{D}_k$ 
7                 recompute  $J_e, \mathbf{m}_i, \mathbf{m}_k$ 
8         until no change in  $J_e$  in  $n$  attempts
9     return  $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_c$ 
10    end
```

Cluster hierárquico

- Considere uma seqüência de particionamentos dos n exemplos em c aglomerados
- O primeiro passo resulta em n aglomerados
- O segundo passo em $n - 1$ aglomerados
- O terceiro passo em $n - 2$ aglomerados
- E assim por diante, até 1 aglomerado

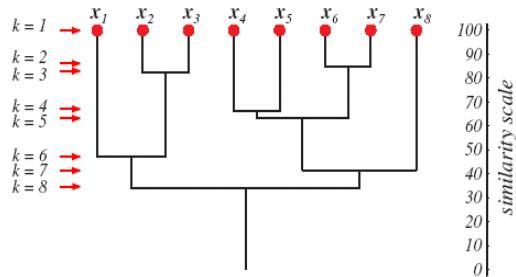
Cluster hierárquico

- Dizemos que estamos no nível k da seqüência se $c=n-k+1$
- No nível 1, $c=n$; no nível n , $c=1$
- Dadas duas amostras x e x' , em algum nível elas vão estar no mesmo aglomerado

Cluster hierárquico

- Se a seqüência tem a propriedade que, quaisquer duas amostras que estão no mesmo aglomerado no nível k , também estejam no mesmo cluster no nível l , $l>k$, então a seqüência é dita ser um cluster hierárquico
- A representação mais natural de um cluster hierárquico é uma árvore chamada dendrograma

Cluster hierárquico



Cluster hierárquico

- Há dois tipos de algoritmos para obter clusters hierárquicos:
 - Aglomerativo – Procedimento bottom-up
 - Neste caso, parte-se de n clusters para, através de um critério de similaridade, chegarmos a apenas um cluster
 - Divisivo – Procedimento top-down
 - Neste caso, parte-se de um cluster com todas as amostras e, através de um critério de separação, divide-se o cluster em dois

Cluster hierárquico

■ Algorithm 4. (Agglomerative Hierarchical Clustering)

```

1 begin initialize  $c, \hat{c} \leftarrow n, \mathcal{D}_i \leftarrow \{\mathbf{x}_i\}, i = 1, \dots, n$ 
2     do  $\hat{c} \leftarrow \hat{c} - 1$ 
3         find nearest clusters, say,  $\mathcal{D}_i$  and  $\mathcal{D}_j$ 
4         merge  $\mathcal{D}_i$  and  $\mathcal{D}_j$ 
5     until  $c = \hat{c}$ 
6     return  $c$  clusters
7 end

```

Cluster hierárquico

- Como medir a distância entre dois clusters?
- Pela mínima distância entre dois pontos quaisquer de cada um dos aglomerados

$$d_{\min}(\Delta_i, \Delta_j) = \min_{\substack{x_i \in \Delta_i \\ x_j \in \Delta_j}} \|x_i - x_j\|$$

Cluster hierárquico

- Pela máxima distância entre dois pontos quaisquer de cada um dos aglomerados

$$d_{\max}(\Delta_i, \Delta_j) = \max_{\substack{x_i \in \Delta_i \\ x_j \in \Delta_j}} \|x_i - x_j\|$$

Cluster hierárquico

- Pela média das distâncias entre todos os pontos dos aglomerados

$$d_{avg}(\Delta_i, \Delta_j) = \frac{1}{n_i n_j} \sum_{x_i \in \Delta_i} \sum_{x_j \in \Delta_j} \|x_i - x_j\|$$

Cluster hierárquico

- Pela distância entre as médias dos dois aglomerados

$$d_{mean}(\Delta_i, \Delta_j) = \|m_i - m_j\|$$

Cluster hierárquico

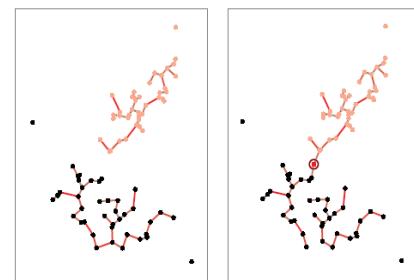
- Pela distância média entre dois pontos quaisquer de cada um dos aglomerados

$$d(\Delta_i, \Delta_j) = \max_{\substack{x_i \in \Delta_i \\ x_j \in \Delta_j}} \|x_i - x_j\|$$

Cluster hierárquico

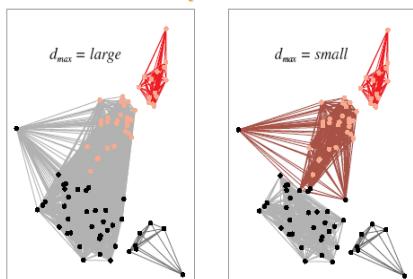
- Quando a distância mínima é usada, o algoritmo de cluster é conhecido como **single-linkage**
- Quando a distância máxima é utilizada, o algoritmo de cluster é conhecido como **complete-linkage**

Cluster hierárquico



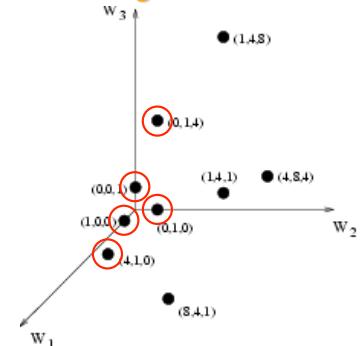
Sensibilidade do single-linkage

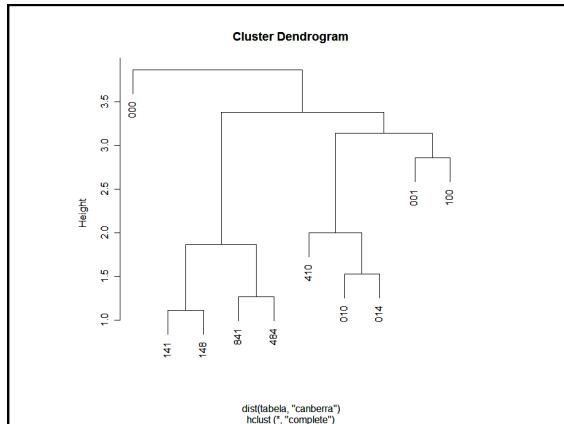
Cluster hierárquico



Sensibilidade do complete-linkage

Análise de aglomerados





Análise de aglomerados

- Retornando ao conjunto de dados, teremos uma nova coluna com os rótulos das regiões

W ₁	W ₂	W ₃	N	Label
0	0	0	14	Fundo
0	0	1	3	Borda
0	1	0	4	Borda
1	0	0	4	Borda
0	1	4	3	Borda
4	1	0	3	Borda
1	4	1	2	Figura
1	4	8	1	Figura
8	4	1	1	Figura
4	8	4	1	Figura

Função critério para aglomerados

- Como medir a qualidade de um cluster?
- Minizando alguma função critério
- Tipos de funções critério:
 - Soma dos erros quadráticos
 - Variância dos erros
 - Matrizes de espalhamento
 - Traço da matriz de espalhamento
 - ...

Soma dos erros quadráticos

- Seja $\Delta = \{x_1, x_2, \dots, x_n\}$ n amostras que queremos partitionar em c subconjuntos disjuntos $\Delta_1, \Delta_2, \dots, \Delta_c$. A soma dos erros quadráticos é dada por:

$$J_e = \sum_{i=1}^c \frac{1}{n_i} \sum_{x \in \Delta_i} \|x - m_i\|^2 \quad m_i = \frac{1}{n_i} \sum_{x \in \Delta_i} x$$

Soma dos erros quadráticos

- A interpretação é simples
- O vetor m_i é o melhor representante das amostras no cluster i pois ele minimiza a soma dos erros quadráticos entre um dos seus elementos e m_i
- O valor J_e é a soma total das contribuições de cada cluster. Minimizase, assim, esse valor