

Curso: Aprendizagem de Máquina em Inteligência Artificial

Disciplina: Aprendizado Não Supervisionado

Prof. Marcelo Novaes de Rezende



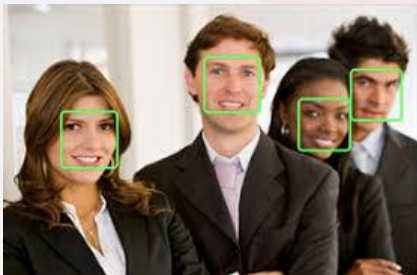
Usando PCA: eigenfaces no reconhecimento facial



Faculdade de Computação e Informática
Mackenzie

Entendendo o Eigenfaces

Face Detection



*São necessárias várias imagens para termos os autovetores do PCA antes de gerar as representações

Face Recognition



Imagem $n \times n$
pixels (n^2
dimensões)



Representação
reduzida $m < n$
(menos
features)



possível comparação com
outros rostos

Faculdade de Computação e Informática
Mackenzie

Entendendo o Eigenfaces

Para começar, vamos partir de um dataset de rostos 64x64:

- a) aplicar PCA (128 dimensões)
- b) gerar as representações de jack1, jack2 e alguém (do dataset)
- c) ver as distâncias (normas) entre jack1 e jack2 e jack1 e alguém
- d) Desafio : mostrar a foto da pessoa mais parecida com jack1 na base...

Eigfaces1.ipynb



Faculdade de Computação e Informática
Mackenzie

■ Regras de Associação

Regras de associação são técnicas para identificar relações entre itens.

Sendo itens $I=\{i_1,i_2,...i_n\}$

Uma regra (há outras definições) é

$X \rightarrow Y$ com $X,Y \subset I$, ou seja, X implica Y com X e Y sendo subconjuntos de I .

Uma transação t é um subconjunto de I

■ Regras de Associação

Métricas usuais em regras de associação

Suporte X (item ou conjunto de itens):

Número de transações em que X aparece/ número total de transações. Exemplo: se café aparece em 50 das 1000 transações diárias de um supermercado, o Suporte(café) é $50/1000=0.05$.

Confiança $X \rightarrow Y$ = $\text{Suporte}(X \cup Y) / \text{Suporte } X$

Exemplo, a compra de café e pão tem suporte 0.02. Assim, a confiança $\{\text{café}\} \rightarrow \{\text{café}, \text{pão}\}$ tem confiança:

$0.02/0.05=0.4(40\%)$...supondo suporte de café=0.05

■ Regras de Associação

Métricas usuais em regras de associação

$$\text{Lift } X \rightarrow Y = \text{Suporte}(X \cup Y) / (\text{Suporte } X * \text{Suporte } Y)$$

Compara o suporte conjunto com os independentes. Se $\text{Lift} > 1$ há provável associação (item “puxa outro”). Se $\text{lift} < 1$, um item “atrapalha o outro”). $\text{Lift}=1$ indica que os itens nem são associados (um puxa outro), nem atrapalham um ao outro na transação.

Outra formulação equivalente:

$$\text{Lift } X \rightarrow Y = \text{Confiância } X \rightarrow Y / \text{suporte de } B \text{ (} X=A, Y=(A,B) \text{)}$$

Indica também o incremento de vendas de B se A for vendido.

■ Regras de Associação

Partindo de rules.ipynb,

Vamos calcular alguns suportes, confianças e lifts

■ Algoritmo Apriori

Apriori é um algoritmo clássico para busca de regras de associação em bases de dados.

A principal ideia é: ***itens infrequentes geram associações infrequentes.***

Partindo dessa ideia principal, colocamos um threshold para suporte de itens individuais, duplas, triplas etc.

Partimos de itens individuais, fazemos a poda (corte por suporte), criamos as duplas + corte por suporte e assim, por diante. Também podemos criar cortes para confiança de regras.

■ Apriori “na mão”

Gerar lista de duplas e triplas e algumas confianças na lista de transações:

```
[['cerveja', 'pão', 'carne'], ['vinho', 'queijo'],  
 ['cerveja', 'carne'], ['vinho', 'pão', 'carne'], ['pão', 'carne'],  
 ['cerveja', 'vinho', 'pão', 'carne'], ['pão', 'vinho'],  
 ['cerveja', 'carne'], ['cerveja', 'carne'], ['vinho', 'pão']]
```

Usaremos threshold 0.2 para suportes

■ Apriori “na mão”

['cerveja', 'pão', 'carne'],
['vinho', 'queijo'],
['cerveja', 'carne'],
['vinho', 'pão', 'carne'],
['pão', 'carne'],
['cerveja', 'vinho', 'pão', 'carne'],
['pão', 'vinho'],
['cerveja', 'carne'],
['cerveja', 'carne'],
['vinho', 'pão']

Poda

Itens individuais: suporte e poda (threshold 0.2)

Item	Contagem	Suporte
pão	6	0,6
cerveja	5	0,5
vinho	5	0,5
carne	7	0,7
queijo	1	0,1

Apriori “na mão”

duplas: suporte e poda (threshold 0.2)

['cerveja', 'pão', 'carne'],

['vinho', 'queijo'],

['cerveja', 'carne'],

['vinho', 'pão', 'carne'],

['pão', 'carne'],

['cerveja', 'vinho', 'pão', 'carne'],

['pão', 'vinho'],

['cerveja', 'carne'],

['cerveja', 'carne'],

['vinho', 'pão']

Item

pão

cerveja

vinho

carne

Poda

Duplas

Count

Suporte

pão	cerveja	2	0,2
pão	vinho	4	0,4
pão	carne	3	0,4
cerveja	vinho	1	0,1
cerveja	carne	5	0,5
vinho	carne	2	0,2

Apriori “na mão”

['cerveja', 'pão', 'carne'],

['vinho', 'queijo'],

['cerveja', 'carne'],

['vinho', 'pão', 'carne'],

['pão', 'carne'],

['cerveja', 'vinho', 'pão', 'carne'],

['pão', 'vinho'],

['cerveja', 'carne'],

['cerveja', 'carne'],

['vinho', 'pão']

Duplas

pão	cerveja
pão	vinho
pão	carne
cerveja	carne
vinho	carne

triplas: suporte e poda (threshold 0.2)

Poda

Triplas			count	Suporte
pão	cerveja	carne	2	0,2
pão	vinho	carne	2	0,2
pão	cerveja	vinho	1	0.1
carne	cerveja	vinho	1	0.1

Apriori “na mão”

Alguns cálculos de confiança e lift nas duplas e triplas

Confiança $\{\text{pão}\} \rightarrow \{\text{pão}, \text{carne}\}$
Suporte($\text{pão}, \text{carne}$)/suporte(pão)
 $0.4/0.6=0.6667$

Confiança $\{\text{vinho}\} \rightarrow \{\text{vinho}, \text{carne}, \text{pão}\}$
Suporte($\text{vinho}, \text{carne}, \text{pão}$)/suporte(vinho)
 $0.2/0.5=0.4$

Lift $\{\text{pão}\} \rightarrow \{\text{pão}, \text{carne}\}$
Suporte($\text{pão}, \text{carne}$)/(suporte(pão)*suporte(carne))
 $0.4/(0.6*0.7)=0.95$

Duplas

pão	cerveja
pão	vinho
pão	carne
cerveja	carne
vinho	carne

Triplas

pão	cerveja	carne
pão	vinho	carne

■ Apriori no Python

Analisar código de **apriori.ipynb**

Conferir resultados dos slides anteriores

É utilizada a biblioteca apyori:

```
from apyori import apriori
transactions = [ ['beer', 'nuts'],
                  ['beer', 'cheese'], ]
results = list(apriori(transactions))
```

■ Apriori no Python - Tarefa

Partir de `apriori_tarefa.ipynb`
Utilizar dataset `store_data.csv`
Gerar a lista de transações (lista de listas)
Definir thresholds (cuidado!)
Rodar apriori
Listar recomendações ao supermercado (promoções
conjuntas ou reposicionamento de gôndolas)

■ Sobrou tempo?

Recomendação Collaborative Filtering

Até a próxima aula

OBRIGADO!

Prof Marcelo Rezende
email rezendemn@gmail.com