

Tópicos Especiais em Sistemas de Informação

**Classificadores de mais de duas categorias
(Multiclassificadores)**

Ely – elydasilvamiranda@gmail.com

Classificações estudadas

- Até o momento classificamos como:
 - É ou não spam;
 - É um porco ou cachorro;
 - Um usuário compra ou não um curso;

Qual a principal característica dessas formas de classificação?

Classificação não binária

- No mundo real, classificações podem ser bem mais complexas;
- Há classificações que possuem mais de duas possibilidades, por exemplo:
 - Qual o **grau de satisfação** de um cliente?
 - Super satisfeito, satisfeito , neutro, insatisfeito, com raiva;
 - Qual o **tipo de e-mail** que chegou?
 - Spam, promoção, fórum, update, importante, familiar ou normal;

Classificação não binária

- A partir da classificação de **grau de satisfação**, podem-se tomar decisões como:
 - Com qual dos clientes a empresa deve interagir?
 - Qual a forma de interação dependendo do grau de satisfação do cliente?
- Com base na classificação de **tipo de e-mail**, podem-se tomar decisões como:
 - Evitar a caixa de entrada;
 - Ignorar o e-mail ou jogar para a lixeira;
 - Criar algum tipo de notificação personalizada.

Estudo de caso: Grau de satisfação

- Algumas características de um determinado cliente:
 - **Último acesso:** Visitou ontem (1 dia atrás);
 - **Frequência de acesso:** Visitou 4 dias;
 - **Se inscreveu:** 4 semanas atrás;
 - **Está se sentindo:** Alegre.

Estudo de caso: Grau de satisfação

- A primeira característica informa a última visita que o cliente fez no site:
 - **Último acesso:** Visitou ontem (1 dia atrás);
 - Essa informação responde a seguinte pergunta: "O quanto recente foi o último acesso desse cliente?".
 - Identificamos isso como **recência**;
 - A recência do cliente ao site será indicada em dias;
- Assim compararemos o que um cliente sente em relação a um outro cliente que acessou há uma semana, ou há um mês ou há ano atrás.

Estudo de caso: Grau de satisfação

- A segunda descreve qual é a frequência de visitas que o cliente teve após sua inscrição:
 - **Frequência de acesso:** Visitou 4 dias;
 - Essa frequência responde à pergunta: O quão frequente foi o acesso do nosso cliente"?;
 - Nesse caso, 4 dias distintos.
 - Esse tipo de informação pode ser relevante:
 - Uma pessoa que acessou 4 dias e outra que acessou nenhuma dia após a inscrição, indica se o produto é ou não interessante.

Estudo de caso: Grau de satisfação

- A próxima característica refere-se ao tempo que o cliente se inscreveu no site:
 - **Se inscreveu:** 4 semanas atrás;
 - Dependendo do tempo de inscrição, cada cliente poderá conter diferentes tipos de comportamentos:
 - Um cliente inscrito semana passada tem um comportamento diferente de um inscrito há 3 anos;
 - Por outro lado, existem situações em que quanto menos tempo estamos atrelados ao produto, mais chateados ficamos.

Estudo de caso: Grau de satisfação

- Para essas 3 variáveis tentaremos prever se o cliente alegre ou neutro ou chateado:
 - **Alegre:**
 - Significa que ele está satisfeito;
 - Não precisamos nos preocupar tanto;
 - **Neutro:**
 - Não demonstra nem satisfação ou insatisfação com o nosso produto;
 - Podemos considerar um contato para entender melhor a situação dele.

Estudo de caso: Grau de satisfação

- Para essas 3 variáveis tentaremos prever se o cliente alegre ou neutro ou chateado:
 - **Chateado:**
 - Está insatisfeito;
 - É de extrema importância entender o que está acontecendo;
 - Precisaremos verificar o motivo e tentar de alguma maneira recuperar a situação.

Modelando o problema

- Arquivo satisfacao_clientes.csv:

| | recencia | frequencia | semanas_de_inscricao | situacao |
|---|----------|------------|----------------------|----------|
| 2 | 1 | 4 | 4 | 2 |
| 3 | 2 | 1 | 2 | 1 |
| 4 | 1 | 4 | 2 | 2 |
| 5 | 1 | 3 | 8 | 1 |
| 6 | 2 | 2 | 1 | 1 |
| 7 | 1 | 4 | 2 | 2 |
| 8 | 1 | 1 | 5 | 1 |
| 9 | 1 | 3 | 8 | 1 |
| 0 | 3 | 1 | 1 | 1 |
| 1 | 3 | 2 | 6 | 1 |
| 2 | 2 | 3 | 6 | 1 |
| 3 | 3 | 1 | 2 | 1 |
| 4 | 3 | 2 | 7 | 1 |
| 5 | 1 | 4 | 2 | 2 |
| 6 | 1 | 3 | 1 | 2 |
| 7 | 1 | 4 | 8 | 1 |
| 8 | 1 | 3 | 7 | 1 |
| 9 | 3 | 1 | 7 | 0 |

| | |
|---|----------|
| 0 | Chateado |
| 1 | Neutro |
| 2 | Alegre |

Implementando

- Como os algoritmos que já conhecemos se comportará com esses dados?
- Será que AdaBoost e Multinomial terão bons resultados?

MultinomialNB e AdaBoost

```
import pandas as pd
import numpy as np
from sklearn.model_selection import
cross_val_score
from scipy.stats import bayes_mvs

df = pd.read_csv('satisfacao_clientes.csv')
x = df[['recencia', 'frequencia',
        'semanas_de_inscricao']]
y = df['situacao'].values
k = 10
```

MultinomialNB e AdaBoost

```
# multinomial
from sklearn.naive_bayes import MultinomialNB
modeloMultinomial = MultinomialNB()

scores = cross_val_score(modeloMultinomial,x,y,cv = k)
print(bayes_mvs(scores,0.95))

Mean(statistic=0.8134466403162055 ...)
```



```
# AdaBoost
from sklearn.ensemble import AdaBoostClassifier
modeloAdaBoost = AdaBoostClassifier()

scores = cross_val_score(modeloAdaBoost,x,y,cv = k)
print(bayes_mvs(scores,0.95))

Mean(statistic=0.7607351778656126 ...)
```

Algoritmos Multiclasses

- Alguns algoritmos que tratam dessa questão reduzindo os conjuntos de valores a apenas dois.
- Assim os resultados possíveis seriam analisados em várias etapas:
 - Resultado: 0 e os demais (1 e 2) seriam 1;
 - Resultado: 1 e os demais (0 e 2) seriam 0;
 - Resultado: 2 e os demais (0 e 1) seriam 0.
- A partir daí, estatísticas seriam extraídas para prever quando cada resultado seria o correto.

Algoritmos Multiclasses

- Dois algoritmos populares são:
 - OneVersusRest;
 - OneVersusOne;
- São algoritmos que também são muito usados para reconhecimento de padrões em I.A.;
- Sua principal característica é o isolamento do conjunto de resultados e execução de vários testes;
- A performance desses algoritmos é inferior aos já vistos até aqui;
- Porém, são os mais indicados para multiclassificação.

OneVsRest e OneVsOne

```
from sklearn.multiclass import OneVsRestClassifier
from sklearn.svm import LinearSVC
modeloOneVsRest = OneVsRestClassifier(LinearSVC(random_state = 0))
```

```
Mean(statistic= 0.9200632411067191 ...)
```

```
from sklearn.multiclass import OneVsOneClassifier
modeloOneVsOne = OneVsOneClassifier(LinearSVC(random_state = 0))
scores = cross_val_score(modeloAdaBoost,x,y,cv = k)
print(scores.mean())
```

1.0

Resultados

| Multinomial | AdaBoost | OneVsRest | OneVsOne |
|-------------|----------|-----------|----------|
| 81,34% | 76,07% | 92% | 100% |

Tópicos Especiais em Sistemas de Informação

**Classificadores de mais de duas categorias
(Multiclassificadores)**

Ely – elydasilvamiranda@gmail.com