



Unyleya
EDUCACIONAL



One hot

Transformando dados em categóricos

Os dados estarem adequados para a mineração de dados é tão importante quanto escolher uma boa abordagem para o problema em questão. Isto porque o estado dos dados pode impactar diretamente na qualidade dos resultados que serão obtidos ao final da execução do algoritmo.

O que são dados categóricos?

Dados categóricos são os chamados nominais ou ordinais.

- Sexo: feminino / masculino;
- Nível de escolaridade: 1º grau, 2º grau, 3º grau.

Como resolver

Alguns algoritmos só conseguem trabalhar bem com dados categóricos, como o SVM. Por isso pode ser muito interessante transformar atributos de um conjunto de dados em categóricos.

Essa transformação pode ser feita de forma bastante simples, utilizando a biblioteca Scikit-learn e um recurso que ela oferece, chamado de Label Encoder.

Como funciona

Se tivermos um atributo do tipo frutas, contendo:

- laranja, banana, uva, carambola.

A transformação dessa variável em categórica resultaria em:

- 0, 1, 2, 3.

Onde:

- laranja = 0; banana = 1; uva = 2; carambola = 3.

Utilizando o labelEncoder

```
import sklearn
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
frutas = ['laranja', 'banana', 'uva', 'carambola']
valoresNumericos = label_encoder.fit_transform(frutas)
```

One Hot Encoding

A diferença do *one hot encoding* é que além de gerar valores categóricos, ele gera representações binárias para esses valores.

```
numeros = [1, 45, 10]
onehot_encoder = OneHotEncoder(sparse=False)
inteiros = valores_numericos.reshape(len(valores_numericos), 1)
valorBinarios = onehot_encoder.fit_transform(inteiros)
```

One Hot Encoding

A diferença do *one hot encoding* é que além de gerar valores categóricos, ele gera representações binárias para esses valores.

```
valorBinarios =  
[1, 0, 0]  
[0, 1, 0]  
[0, 0, 1]
```




Obrigada!

Ana Laurentino