

04

Para saber mais

Nesse curso aprendemos a utilizar o método LabelEncoder do Scikit Learn para converter variáveis categóricas em uma representação numérica. No entanto, existem outros métodos para tratar esse tipo de variável, um dos mais conhecidos é o Dummies Variables.

Fazer a transformação de uma variável categórica para uma variável dummy é tão simples quanto o nome já sugere(dummy pode ser traduzido como bobo em português). Cada valor único da variável se transforma em uma coluna booleana(Verdadeiro ou Falso).

Suponhamos que temos uma cola com a informação do sexo de algumas pessoas.

sex
male
female
male
female
female

Ao criar variáveis dummies a partir dessa coluna, teríamos duas novas colunas, caso o valor original fosse “male” teremos 1 na coluna “male” e 0 na “female”, caso fosse “female”, teríamos o oposto, 0 em “male” e 1 em “female”.

female	male
0.0	1.0
1.0	0.0
0.0	1.0
1.0	0.0
1.0	0.0

O que facilita bastante a utilização de dummies é que o pandas já vem com uma função que pode fazer a conversão de forma bem simples:

```
pandas.get_dummies(dataframe)
```