



Normalização de dados

A normalização

A normalização é importante para criar uma padronização do comportamento dos dados que serão utilizados, facilitando o aprendizado correto do algoritmo.

Com a normalização, os dados numéricos serão dimensionados para um intervalo específico, que pode ser -1 a 1 ou de 0 a 1.

Essa ação pode ser necessária em algumas das bases que forem utilizadas na construção de abordagens com o aprendizado de máquina.

Formas mais utilizadas

Existem várias formas de normalização, então serão apresentadas apenas duas das mais utilizadas, que são:

- método min-max;
- método z-score.

Normalização - Min Max

A normalização Min-Max utiliza um valor mínimo e máximo para determinar o intervalo para qual os valores serão transportados. Considerando o atributo **A**, cada ocorrência de **v** é transformada em **v'**, dentro da faixa [novoMin, novoMax].

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{novoMax} - \text{novoMin}) + \text{novoMin}$$

Utilizando o Python - Min Max

Essa normalização pode ser realizada facilmente através da linguagem Python. Para isso devem ser importadas as seguintes bibliotecas.

```
import pandas as pd  
import numpy as np  
from sklearn.preprocessing import MinMaxScaler
```

Utilizando o Python - Min Max

A seguir são definidas as colunas e é criado um DataFrame com os dados que serão normalizados.

```
#Definindo colunas
```

```
cols =['B', 'C']
```

```
#Carregando dados
```

```
data = pd.DataFrame(np.array([[2, 3],  
[1.02, 1.2],  
[0.5, 0.3]]), columns=cols)
```

Utilizando o Python - Min Max

Foi definida uma escala para ser aplicada de 0 a 1, em seguida os dados são transformados.

```
#Definido o Min e Max
```

```
scaler = MinMaxScaler(feature_range = (0, 1))
```

```
#Transformando os dados
```

```
transformedData = scaler.fit_transform(data[cols])
```

```
#Exibindo os resultados
```

```
print(transformedData)
```

Normalização - z-score

A normalização z-score utiliza a conversão de valores utilizando a média e o desvio padrão do atributo. A equação que define esse método é:

$$v' = \frac{v - \underline{A}}{\sigma_A}$$

Utilizando o Python - z-score

A normalização z-score pode ser executada de forma simples no Python.

```
import numpy as np

from scipy import stats

arr1 = [[20, 2, 7, 1, 34],
        [50, 12, 12, 34, 4]]

resultado = stats.zscore(arr1)
```



Obrigada!

Ana Laurentino