5_feature_scaling_exercise

May 20, 2019

In [ ]:
```
# code from - https://stackoverflow.com/questions/24645153/pandas-dataframe-columns-sc

import pandas as pd
import numpy as np

df = pd.DataFrame({'A':[14.00,90.20,90.95,96.27,91.21],'B':[103.02,107.26,\
        110.35,114.23,114.68], 'C':['big','small','big','small','small']})
df
```

In [ ]:
```
df["A"]
```

In [ ]:
```
df["A"]  - df["A"].min()
```

In [ ]:
```
( df["A"] - df["A"].min() ) / (df["A"].max() - df["A"].min())
```

In [ ]:
```
df["A"] = ( df["A"] - df["A"].min() )  \
/ (df["A"].max() - df["A"].min()) * (5 - 1) + 1
df
```

In [ ]:
```
df["B"].mean(), df["B"].std()
```

In [ ]:
```
df["B"] = ( df["B"] - df["B"].mean() )  \
/ (df["B"].std() )
```

In [ ]:
```
df
```

In [ ]:
```
def feture_scaling(df, scaling_strategy="min-max", column=None):
    if column == None:
        column = [column_name for column_name in df.columns]
    for column_name in column:
        if scaling_strategy == "min-max":
            df[column_name] = ( df[column_name] - df[column_name].min() ) /\
                              (df[column_name].max() - df[column_name].min())
        elif scaling_strategy == "z-score":
            df[column_name] = ( df[column_name] - \
                                df[column_name].mean() ) /\
                              (df[column_name].std() )
    return df
```

```
In [ ]: df = pd.DataFrame({'A':[14.00,90.20,90.95,96.27,91.21],'B':[103.02,107.26,\
                110.35,114.23,114.68], 'C':['big','small','big','small','small']})
        df

In [ ]: feture_scaling(df,column=["A","B"])

In [ ]: # code from - http://sebastianraschka.com/Articles/2014_about_feature_scaling.html

        import pandas as pd
        import numpy as np

        df = pd.io.parsers.read_csv(
            'https://raw.githubusercontent.com/rasbt/pattern_classification/\
            master/data/wine_data.csv',
             header=None,
             usecols=[0,1,2]
            )

        df.columns=['Class label', 'Alcohol', 'Malic acid']

        df.head()

In [ ]: df = feture_scaling(df, "min-max", column=['Alcohol', 'Malic acid'])
        df.head()

In [ ]: from sklearn import preprocessing

        df = pd.io.parsers.read_csv(
            'https://raw.githubusercontent.com/rasbt/pattern_classification/\
            master/data/wine_data.csv',
             header=None,
             usecols=[0,1,2]
            )
        df.columns=['Class label', 'Alcohol', 'Malic acid']
        df

In [ ]: from sklearn import preprocessing
        std_scaler = preprocessing.StandardScaler().fit(df[['Alcohol', 'Malic acid']])

In [ ]: df_std = std_scaler.transform(df[['Alcohol', 'Malic acid']])

In [ ]: df_std

In [ ]: minmax_scaler = preprocessing.MinMaxScaler().fit(df[['Alcohol', 'Malic acid']])

In [ ]: minmax_scaler.transform(df[['Alcohol', 'Malic acid']])

In [ ]:
```