# lab4-datamining-211220051

April 29, 2024

[ ]:
```
# MIN-MAX SCALING
import pandas as pd
import numpy as np

df = pd.read_csv("/content/train.csv")

def min_max_scaling(df):
    # copy the dataframe
    df_norm = df.copy()
    # apply min-max scaling
    for column in df_norm.select_dtypes(include=np.number).columns:
        df_norm[column] = (df_norm[column] - df_norm[column].min()) /␣
 ↪(df_norm[column].max() - df_norm[column].min())

    return df_norm

# call the min_max_scaling function
df_cars_normalized = min_max_scaling(df)

df_cars_normalized
```

[ ]:
```
     PassengerId  Survived  Pclass  \
0       0.000000       0.0     1.0
1       0.001124       1.0     0.0
2       0.002247       1.0     1.0
3       0.003371       1.0     0.0
4       0.004494       0.0     1.0
..           …       …     …
886     0.995506       0.0     0.5
887     0.996629       1.0     0.0
888     0.997753       0.0     1.0
889     0.998876       1.0     0.0
890     1.000000       0.0     1.0

                                                 Name     Sex       Age  \
0                             Braund, Mr. Owen Harris    male  0.271174
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  0.472229
```

```
2                                Heikkinen, Miss. Laina  female  0.321438
3          Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  0.434531
4                              Allen, Mr. William Henry    male  0.434531
..                                                  …     …       …
886                              Montvila, Rev. Juozas    male  0.334004
887                       Graham, Miss. Margaret Edith  female  0.233476
888            Johnston, Miss. Catherine Helen "Carrie"  female       NaN
889                              Behr, Mr. Karl Howell    male  0.321438
890                                Dooley, Mr. Patrick    male  0.396833

     SibSp     Parch            Ticket      Fare Cabin Embarked
0    0.125  0.000000         A/5 21171  0.014151   NaN        S
1    0.125  0.000000          PC 17599  0.139136   C85        C
2    0.000  0.000000  STON/O2. 3101282  0.015469   NaN        S
3    0.125  0.000000            113803  0.103644  C123        S
4    0.000  0.000000            373450  0.015713   NaN        S
..     …       …                 …        …   …       …
886  0.000  0.000000            211536  0.025374   NaN        S
887  0.000  0.000000            112053  0.058556   B42        S
888  0.125  0.333333        W./C. 6607  0.045771   NaN        S
889  0.000  0.000000            111369  0.058556  C148        C
890  0.000  0.000000            370376  0.015127   NaN        Q

[891 rows x 12 columns]
```

[ ]:
```python
# Z-SCORE
def z_score(df):
    # copy the dataframe
    df_std = df.copy()
    # apply the z-score method
    for column in df_std.select_dtypes(include=np.number).columns:
        df_std[column] = (df_std[column] - df_std[column].mean()) /␣
 ↪df_std[column].std()

    return df_std

# call the z_score function
df_cars_standardized = z_score(df)

df_cars_standardized
```

[ ]:
```
     PassengerId  Survived    Pclass  \
0      -1.729137 -0.788829  0.826913
1      -1.725251  1.266279 -1.565228
2      -1.721365  1.266279  0.826913
3      -1.717480  1.266279 -1.565228
4      -1.713594 -0.788829  0.826913
```

```
..           …       …       …
886     1.713594 -0.788829 -0.369158
887     1.717480  1.266279 -1.565228
888     1.721365 -0.788829  0.826913
889     1.725251  1.266279 -1.565228
890     1.729137 -0.788829  0.826913

                                                  Name     Sex       Age  \
0                              Braund, Mr. Owen Harris    male -0.530005
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  0.571430
2                               Heikkinen, Miss. Laina  female -0.254646
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  0.364911
4                             Allen, Mr. William Henry    male  0.364911
..                                                 …     …       …
886                             Montvila, Rev. Juozas    male -0.185807
887                      Graham, Miss. Margaret Edith  female -0.736524
888           Johnston, Miss. Catherine Helen "Carrie"  female       NaN
889                              Behr, Mr. Karl Howell    male -0.254646
890                                Dooley, Mr. Patrick    male  0.158392

        SibSp     Parch            Ticket      Fare Cabin Embarked
0    0.432550 -0.473408         A/5 21171 -0.502163   NaN        S
1    0.432550 -0.473408          PC 17599  0.786404   C85        C
2   -0.474279 -0.473408  STON/O2. 3101282 -0.488580   NaN        S
3    0.432550 -0.473408            113803  0.420494  C123        S
4   -0.474279 -0.473408            373450 -0.486064   NaN        S
..        …       …                 …       …   …      …
886 -0.474279 -0.473408            211536 -0.386454   NaN        S
887 -0.474279 -0.473408            112053 -0.044356   B42        S
888  0.432550  2.007806        W./C. 6607 -0.176164   NaN        S
889 -0.474279 -0.473408            111369 -0.044356  C148        C
890 -0.474279 -0.473408            370376 -0.492101   NaN        Q

[891 rows x 12 columns]
```

[ ]:
```python
# MAXIMUM ABSOLUTE SCALING
def maximum_absolute_scaling(df):
    # copy the dataframe
    df_scaled = df.copy()
    # apply maximum absolute scaling
    for column in df_scaled.select_dtypes(include=np.number).columns:
        df_scaled[column] = df_scaled[column]  / df_scaled[column].abs().max()
    return df_scaled

# call the maximum_absolute_scaling function
df_cars_scaled = maximum_absolute_scaling(df)
```

```
df_cars_scaled
```

[ ]:
```
     PassengerId  Survived    Pclass  \
0       0.001122       0.0  1.000000
1       0.002245       1.0  0.333333
2       0.003367       1.0  1.000000
3       0.004489       1.0  0.333333
4       0.005612       0.0  1.000000
..           …       …       …
886     0.995511       0.0  0.666667
887     0.996633       1.0  0.333333
888     0.997755       0.0  1.000000
889     0.998878       1.0  0.333333
890     1.000000       0.0  1.000000

                                                  Name     Sex     Age  SibSp  \
0                              Braund, Mr. Owen Harris    male  0.2750  0.125
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  0.4750  0.125
2                               Heikkinen, Miss. Laina  female  0.3250  0.000
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  0.4375  0.125
4                             Allen, Mr. William Henry    male  0.4375  0.000
..                                                 …     …     …    …
886                             Montvila, Rev. Juozas    male  0.3375  0.000
887                      Graham, Miss. Margaret Edith  female  0.2375  0.000
888          Johnston, Miss. Catherine Helen "Carrie"  female     NaN  0.125
889                             Behr, Mr. Karl Howell    male  0.3250  0.000
890                               Dooley, Mr. Patrick    male  0.4000  0.000

        Parch            Ticket      Fare Cabin Embarked
0    0.000000         A/5 21171  0.014151   NaN        S
1    0.000000          PC 17599  0.139136   C85        C
2    0.000000  STON/O2. 3101282  0.015469   NaN        S
3    0.000000            113803  0.103644  C123        S
4    0.000000            373450  0.015713   NaN        S
..        …                 …       …   …      …
886  0.000000            211536  0.025374   NaN        S
887  0.000000            112053  0.058556   B42        S
888  0.333333        W./C. 6607  0.045771   NaN        S
889  0.000000            111369  0.058556  C148        C
890  0.000000            370376  0.015127   NaN        Q

[891 rows x 12 columns]
```

[ ]:
```
# ROBUST SCALING
def robust_scaling(df):
    # copy the dataframe
    df_robust = df.copy()
```

```
    # apply robust scaling
    for column in df_robust.select_dtypes(include=np.number).columns:
        df_robust[column] = (df_robust[column] - df_robust[column].median())  /␣
 ↪(df_robust[column].quantile(0.75) - df_robust[column].quantile(0.25))
    return df_robust

# call the robust_scaling function
df_cars_robust = robust_scaling(df)

df_cars_robust
```

[ ]:

| | PassengerId | Survived | Pclass | Name | Sex | Age | SibSp | Parch | Ticket | Fare | Cabin | Embarked |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | -1.000000 | 0.0 | 0.0 | Braund, Mr. Owen Harris | male | -0.335664 | 1.0 | NaN | A/5 21171 | -0.312011 | NaN | S |
| 1 | -0.997753 | 1.0 | -2.0 | Cumings, Mrs. John Bradley (Florence Briggs Th… | female | 0.559441 | 1.0 | NaN | PC 17599 | 2.461242 | C85 | C |
| 2 | -0.995506 | 1.0 | 0.0 | Heikkinen, Miss. Laina | female | -0.111888 | 0.0 | NaN | STON/O2. 3101282 | -0.282777 | NaN | S |
| 3 | -0.993258 | 1.0 | -2.0 | Futrelle, Mrs. Jacques Heath (Lily May Peel) | female | 0.391608 | 1.0 | NaN | 113803 | 1.673732 | C123 | S |
| 4 | -0.991011 | 0.0 | 0.0 | Allen, Mr. William Henry | male | 0.391608 | 0.0 | NaN | 373450 | -0.277363 | NaN | S |
| .. | … | … | … | … | … | … | … | … | … | … | … | … |
| 886 | 0.991011 | 0.0 | -1.0 | Montvila, Rev. Juozas | male | -0.055944 | 0.0 | NaN | 211536 | -0.062981 | NaN | S |
| 887 | 0.993258 | 1.0 | -2.0 | Graham, Miss. Margaret Edith | female | -0.503497 | 0.0 | NaN | 112053 | 0.673281 | B42 | S |
| 888 | 0.995506 | 0.0 | 0.0 | Johnston, Miss. Catherine Helen "Carrie" | female | NaN | 1.0 | inf | W./C. 6607 | 0.389604 | NaN | S |
| 889 | 0.997753 | 1.0 | -2.0 | Behr, Mr. Karl Howell | male | -0.111888 | | | | | | |
| 890 | 1.000000 | 0.0 | 0.0 | Dooley, Mr. Patrick | male | 0.223776 | | | | | | |

```
889    0.0    NaN            111369  0.673281  C148        C
890    0.0    NaN            370376 -0.290356   NaN        Q

[891 rows x 12 columns]
```

[ ]:
```
# MEAN NORMALISATION
def mean_normalisation(df):
    # copy the dataframe
    df_mean = df.copy()
    # apply the mean_normalisation method
    for column in df_mean.select_dtypes(include=np.number).columns:
        df_mean[column] = (df_mean[column] - df_mean[column].mean()) /␣
 ↪(df_mean[column].max() - df_mean[column].min())

    return df_mean

# call the mean_normalisation function
df_cars_standardized = mean_normalisation(df)

df_cars_standardized
```

[ ]:
```
     PassengerId  Survived    Pclass  \
0      -0.500000 -0.383838  0.345679
1      -0.498876  0.616162 -0.654321
2      -0.497753  0.616162  0.345679
3      -0.496629  0.616162 -0.654321
4      -0.495506 -0.383838  0.345679
..           …       …       …
886     0.495506 -0.383838 -0.154321
887     0.496629  0.616162 -0.654321
888     0.497753 -0.383838  0.345679
889     0.498876  0.616162 -0.654321
890     0.500000 -0.383838  0.345679

                                                 Name     Sex       Age  \
0                             Braund, Mr. Owen Harris    male -0.096747
1    Cumings, Mrs. John Bradley (Florence Briggs Th…  female  0.104309
2                               Heikkinen, Miss. Laina  female -0.046483
3         Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  0.066611
4                             Allen, Mr. William Henry    male  0.066611
..                                                 …     …       …
886                             Montvila, Rev. Juozas    male -0.033917
887                      Graham, Miss. Margaret Edith  female -0.134445
888          Johnston, Miss. Catherine Helen "Carrie"  female       NaN
889                             Behr, Mr. Karl Howell    male -0.046483
890                               Dooley, Mr. Patrick    male  0.028913
```

```
        SibSp     Parch            Ticket      Fare Cabin Embarked
0    0.059624 -0.063599         A/5 21171 -0.048707   NaN        S
1    0.059624 -0.063599          PC 17599  0.076277   C85        C
2   -0.065376 -0.063599  STON/O2. 3101282 -0.047390   NaN        S
3    0.059624 -0.063599            113803  0.040786  C123        S
4   -0.065376 -0.063599            373450 -0.047146   NaN        S
..        …       …                …       …   …      …
886 -0.065376 -0.063599            211536 -0.037484   NaN        S
887 -0.065376 -0.063599            112053 -0.004302   B42        S
888  0.059624  0.269734        W./C. 6607 -0.017087   NaN        S
889 -0.065376 -0.063599            111369 -0.004302  C148        C
890 -0.065376 -0.063599            370376 -0.047731   NaN        Q

[891 rows x 12 columns]
```

[ ]:
```
import pygwalker as pyg

walker = pyg.walk(df_cars_standardized)
```

```
WARNING: parse invoke code failed, This may affect feature of export code.
WARNING:pygwalker.services.format_invoke_walk_code:parse invoke code failed,
This may affect feature of export code.
```

```
Box(children=(HTML(value='<div id="ifr-pyg-0006109f800d4105QtzV2yfPO5o7TK3H"␣
 ↪style="height: auto">\n    <head>…
```

```
<IPython.core.display.HTML object>
```