STANDARDIZATION AND NORMALIZATION

June 30, 2022

```
[ ]: # NORMALIZATION FORMULA = X-Xmin/Xmax-Xmin
     # LOOK THROUGH THE NOTES
```

```
[4]: import numpy as np
     import pandas as pd
     import seaborn as sns
```

```
[ ]: #to understand normalization first we should look at the data in the csv file
```

```
[5]: df = pd.read_csv('wine.csv')
     print(df)
```

```
     Wine  Alcohol  Malic.acid   Ash   Acl   Mg  Phenols  Flavanoids  \
0       1    14.23        1.71  2.43  15.6  127     2.80        3.06
1       1    13.20        1.78  2.14  11.2  100     2.65        2.76
2       1    13.16        2.36  2.67  18.6  101     2.80        3.24
3       1    14.37        1.95  2.50  16.8  113     3.85        3.49
4       1    13.24        2.59  2.87  21.0  118     2.80        2.69
..    ...      ...         ...   ...   ...  ...      ...         ...
173     3    13.71        5.65  2.45  20.5   95     1.68        0.61
174     3    13.40        3.91  2.48  23.0  102     1.80        0.75
175     3    13.27        4.28  2.26  20.0  120     1.59        0.69
176     3    13.17        2.59  2.37  20.0  120     1.65        0.68
177     3    14.13        4.10  2.74  24.5   96     2.05        0.76

     Nonflavanoid.phenols  Proanth  Color.int   Hue    OD  Proline
0                    0.28     2.29       5.64  1.04  3.92     1065
1                    0.26     1.28       4.38  1.05  3.40     1050
2                    0.30     2.81       5.68  1.03  3.17     1185
3                    0.24     2.18       7.80  0.86  3.45     1480
4                    0.39     1.82       4.32  1.04  2.93      735
..                    ...      ...        ...   ...   ...      ...
173                  0.52     1.06       7.70  0.64  1.74      740
174                  0.43     1.41       7.30  0.70  1.56      750
175                  0.43     1.35      10.20  0.59  1.56      835
176                  0.53     1.46       9.30  0.60  1.62      840
177                  0.56     1.35       9.20  0.61  1.60      560
```

```
[178 rows x 14 columns]
```

[6]: `df.head()`

[6]:
```
   Wine  Alcohol  Malic.acid   Ash   Acl   Mg  Phenols  Flavanoids  \
0     1    14.23        1.71  2.43  15.6  127     2.80        3.06
1     1    13.20        1.78  2.14  11.2  100     2.65        2.76
2     1    13.16        2.36  2.67  18.6  101     2.80        3.24
3     1    14.37        1.95  2.50  16.8  113     3.85        3.49
4     1    13.24        2.59  2.87  21.0  118     2.80        2.69

   Nonflavanoid.phenols  Proanth  Color.int   Hue    OD  Proline
0                  0.28     2.29       5.64  1.04  3.92     1065
1                  0.26     1.28       4.38  1.05  3.40     1050
2                  0.30     2.81       5.68  1.03  3.17     1185
3                  0.24     2.18       7.80  0.86  3.45     1480
4                  0.39     1.82       4.32  1.04  2.93      735
```

[21]:
```python
#intially lets just read the first 3 columns and rename it
df = pd.read_csv('wine.csv',usecols = [0,1,2])
df.head()
```

[21]:
```
   Wine  Alcohol  Malic.acid
0     1    14.23        1.71
1     1    13.20        1.78
2     1    13.16        2.36
3     1    14.37        1.95
4     1    13.24        2.59
```

[10]: `!pip install scikit-learn`

```
Requirement already satisfied: scikit-learn in
c:\users\91936\anaconda3\lib\site-packages (0.24.1)
Requirement already satisfied: joblib>=0.11 in
c:\users\91936\anaconda3\lib\site-packages (from scikit-learn) (1.0.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\91936\anaconda3\lib\site-packages (from scikit-learn) (2.1.0)
Requirement already satisfied: scipy>=0.19.1 in
c:\users\91936\anaconda3\lib\site-packages (from scikit-learn) (1.6.2)
Requirement already satisfied: numpy>=1.13.3 in
c:\users\91936\anaconda3\lib\site-packages (from scikit-learn) (1.20.1)
```

[22]:
```python
import sklearn as sl
```

[23]:
```python
#to perform normalization we have to import few things
from sklearn.preprocessing import MinMaxScaler
```

```
[24]: scaler = MinMaxScaler()
      scaler.fit_transform(df[['Alcohol','Malic.acid']])
      #therefore it has scaled down the feature between the values of 0-1
```

```
[24]: array([[0.84210526, 0.1916996 ],
             [0.57105263, 0.2055336 ],
             [0.56052632, 0.3201581 ],
             [0.87894737, 0.23913043],
             [0.58157895, 0.36561265],
             [0.83421053, 0.20158103],
             [0.88421053, 0.22332016],
             [0.79736842, 0.27865613],
             [1.        , 0.17786561],
             [0.74473684, 0.12055336],
             [0.80789474, 0.28063241],
             [0.81315789, 0.14624506],
             [0.71578947, 0.19565217],
             [0.97894737, 0.19565217],
             [0.88157895, 0.22332016],
             [0.68421053, 0.21146245],
             [0.86052632, 0.23320158],
             [0.73684211, 0.16403162],
             [0.83157895, 0.16798419],
             [0.68684211, 0.46640316],
             [0.79736842, 0.17588933],
             [0.5       , 0.60474308],
             [0.70526316, 0.22134387],
             [0.47894737, 0.16996047],
             [0.65      , 0.21146245],
             [0.53157895, 0.25889328],
             [0.62105263, 0.20355731],
             [0.59736842, 0.19367589],
             [0.74736842, 0.22924901],
             [0.78684211, 0.18577075],
             [0.71052632, 0.15019763],
             [0.67105263, 0.18181818],
             [0.69736842, 0.21541502],
             [0.71842105, 0.15612648],
             [0.65263158, 0.20948617],
             [0.64473684, 0.21146245],
             [0.59210526, 0.17786561],
             [0.53157895, 0.1798419 ],
             [0.53684211, 0.15019763],
             [0.83947368, 0.64229249],
             [0.66578947, 0.1916996 ],
             [0.62631579, 0.61264822],
             [0.75      , 0.22727273],
```

```
[0.58157895, 0.64031621],
[0.53157895, 0.20355731],
[0.83684211, 0.65217391],
[0.88157895, 0.56324111],
[0.75526316, 0.18577075],
[0.80789474, 0.25296443],
[0.76578947, 0.19565217],
[0.53157895, 0.19565217],
[0.73684211, 0.1798419 ],
[0.73421053, 0.19960474],
[0.72105263, 0.22924901],
[0.71315789, 0.18379447],
[0.66578947, 0.19565217],
[0.83947368, 0.18972332],
[0.59473684, 0.243083  ],
[0.70789474, 0.13636364],
[0.35263158, 0.03952569],
[0.34210526, 0.07114625],
[0.42368421, 0.12252964],
[0.69473684, 0.10079051],
[0.35263158, 0.0770751 ],
[0.3       , 0.14031621],
[0.35263158, 0.09288538],
[0.54736842, 0.05335968],
[0.35263158, 0.08498024],
[0.60789474, 0.03952569],
[0.31052632, 0.08893281],
[0.33157895, 0.17193676],
[0.74473684, 0.15217391],
[0.64736842, 0.18181818],
[0.51578947, 0.18379447],
[0.24473684, 0.06916996],
[0.16578947, 0.22529644],
[0.52631579, 0.03162055],
[0.21315789, 0.42490119],
[0.34210526, 0.04940711],
[0.43947368, 0.61857708],
[0.25526316, 0.03557312],
[0.44473684, 0.21146245],
[0.27631579, 0.0770751 ],
[0.53157895, 0.61660079],
[0.21315789, 0.02964427],
[0.43157895, 0.04743083],
[0.29736842, 0.17193676],
[0.16315789, 0.18379447],
[0.16052632, 0.26086957],
[0.27631579, 0.11660079],
```

```
[0.27631579, 0.21541502],
[0.25526316, 0.15217391],
[0.43684211, 0.15612648],
[0.33157895, 0.41304348],
[0.15526316, 0.24703557],
[0.37894737, 0.1541502 ],
[0.20526316, 0.27272727],
[0.33157895, 0.13241107],
[0.35263158, 0.06521739],
[0.33157895, 0.48023715],
[0.27631579, 0.26482213],
[0.41315789, 0.11857708],
[0.34473684, 0.33794466],
[0.20789474, 0.19367589],
[0.38947368, 0.19565217],
[0.36578947, 0.35770751],
[0.32105263, 0.19565217],
[0.44473684, 0.19960474],
[0.31315789, 0.10869565],
[0.15263158, 0.12055336],
[0.11315789, 0.59288538],
[0.39210526, 0.33399209],
[0.19210526, 0.38339921],
[0.1       , 0.        ],
[0.27631579, 0.1284585 ],
[0.        , 0.15217391],
[0.20789474, 0.14426877],
[0.36578947, 0.17193676],
[0.45789474, 0.53162055],
[0.25526316, 0.53162055],
[0.11052632, 0.32806324],
[0.13947368, 0.25889328],
[0.36578947, 0.72924901],
[0.53157895, 1.        ],
[0.22105263, 0.7055336 ],
[0.27368421, 0.28063241],
[0.36842105, 0.15612648],
[0.2       , 0.27470356],
[0.35263158, 0.17588933],
[0.26578947, 0.70355731],
[0.48157895, 0.12055336],
[0.48684211, 0.44466403],
[0.46842105, 0.31027668],
[0.43947368, 0.55533597],
[0.38947368, 0.09881423],
[0.41315789, 0.33992095],
[0.32105263, 0.78656126],
```

```
       [0.39473684, 0.94268775],
       [0.64736842, 0.56324111],
       [0.47631579, 0.43873518],
       [0.5       , 0.40909091],
       [0.61315789, 0.35968379],
       [0.65526316, 0.48023715],
       [0.68157895, 0.83201581],
       [0.32105263, 0.62055336],
       [0.56052632, 0.55928854],
       [0.75      , 0.84980237],
       [0.48421053, 0.76482213],
       [0.60263158, 0.49407115],
       [0.53947368, 0.62450593],
       [0.65      , 0.47035573],
       [0.46315789, 0.38142292],
       [0.54736842, 0.22924901],
       [0.57894737, 0.50592885],
       [0.40789474, 0.10869565],
       [0.56315789, 0.87944664],
       [0.73947368, 0.66798419],
       [0.37368421, 0.45256917],
       [0.87105263, 0.18577075],
       [0.64473684, 0.18379447],
       [0.35      , 0.61067194],
       [0.7       , 0.49802372],
       [0.47894737, 0.5       ],
       [0.50789474, 0.53557312],
       [0.72368421, 0.39920949],
       [0.71052632, 0.71541502],
       [0.63684211, 0.58498024],
       [0.47105263, 0.51976285],
       [0.67105263, 0.36363636],
       [0.62368421, 0.76284585],
       [0.30789474, 0.45256917],
       [0.45789474, 0.32608696],
       [0.82368421, 0.34980237],
       [0.70526316, 0.97035573],
       [0.62368421, 0.62648221],
       [0.58947368, 0.69960474],
       [0.56315789, 0.36561265],
       [0.81578947, 0.66403162]])
```

[ ]: #to understand standardization.
     #   STANDARDIZATION(Z-SCORE NORMALIZATION)

[ ]: #here all the features will be transformed in such a way that it will have mean␣
     ↪value = 0, std dev = 1

```
#formua = x - mean / std dev
```

[25]:
```
from sklearn.preprocessing import StandardScaler
```

[26]:
```
scaling = StandardScaler()
scaling.fit_transform(df[['Alcohol','Malic.acid']])
```

[26]:
```
array([[ 1.51861254, -0.5622498 ],
       [ 0.24628963, -0.49941338],
       [ 0.19687903,  0.02123125],
       [ 1.69154964, -0.34681064],
       [ 0.29570023,  0.22769377],
       [ 1.48155459, -0.51736664],
       [ 1.71625494, -0.4186237 ],
       [ 1.3086175 , -0.16727801],
       [ 2.25977152, -0.62508622],
       [ 1.0615645 , -0.88540853],
       [ 1.3580281 , -0.15830138],
       [ 1.38273339, -0.76871232],
       [ 0.92568536, -0.54429654],
       [ 2.16095032, -0.54429654],
       [ 1.70390229, -0.4186237 ],
       [ 0.77745356, -0.47248348],
       [ 1.60508109, -0.37374054],
       [ 1.02450655, -0.68792264],
       [ 1.46920194, -0.66996938],
       [ 0.78980621,  0.68550197],
       [ 1.3086175 , -0.63406285],
       [-0.08723191,  1.31386618],
       [ 0.87627476, -0.42760033],
       [-0.18605311, -0.66099274],
       [ 0.61686912, -0.47248348],
       [ 0.06099988, -0.25704433],
       [ 0.48098997, -0.50839001],
       [ 0.36981612, -0.55327317],
       [ 1.07391715, -0.3916938 ],
       [ 1.2592069 , -0.58917969],
       [ 0.90098006, -0.75075906],
       [ 0.71569031, -0.60713296],
       [ 0.83921681, -0.45453022],
       [ 0.93803801, -0.72382916],
       [ 0.62922177, -0.48146012],
       [ 0.59216382, -0.47248348],
       [ 0.34511082, -0.62508622],
       [ 0.06099988, -0.61610959],
       [ 0.08570518, -0.75075906],
       [ 1.50625989,  1.48442217],
```

```
[ 0.69098501, -0.5622498 ],
[ 0.50569527,  1.3497727 ],
[ 1.0862698 , -0.40067043],
[ 0.29570023,  1.47544554],
[ 0.06099988, -0.50839001],
[ 1.49390724,  1.52930533],
[ 1.70390229,  1.12535692],
[ 1.1109751 , -0.58917969],
[ 1.3580281 , -0.28397422],
[ 1.1603857 , -0.54429654],
[ 0.06099988, -0.54429654],
[ 1.02450655, -0.61610959],
[ 1.01215391, -0.52634327],
[ 0.95039066, -0.3916938 ],
[ 0.91333271, -0.59815632],
[ 0.69098501, -0.54429654],
[ 1.50625989, -0.57122643],
[ 0.35746347, -0.32885738],
[ 0.88862741, -0.81359548],
[-0.77898029, -1.25345042],
[-0.82839089, -1.10982432],
[-0.44545875, -0.8764319 ],
[ 0.82686416, -0.97517485],
[-0.77898029, -1.08289442],
[-1.02603329, -0.79564222],
[-0.77898029, -1.01108137],
[ 0.13511578, -1.190614  ],
[-0.77898029, -1.0469879 ],
[ 0.41922672, -1.25345042],
[-0.97662269, -1.02903463],
[-0.87780149, -0.65201611],
[ 1.0615645 , -0.74178243],
[ 0.60451647, -0.60713296],
[-0.01311602, -0.59815632],
[-1.28543893, -1.11880095],
[-1.65601842, -0.40964706],
[ 0.03629458, -1.28935695],
[-1.43367073,  0.49699271],
[-0.82839089, -1.20856726],
[-0.37134286,  1.3767026 ],
[-1.23602833, -1.27140368],
[-0.34663756, -0.47248348],
[-1.13720713, -1.08289442],
[ 0.06099988,  1.36772596],
[-1.43367073, -1.29833358],
[-0.4084008 , -1.21754389],
[-1.03838594, -0.65201611],
```

```
[-1.66837107, -0.59815632],
[-1.68072372, -0.2480677 ],
[-1.13720713, -0.90336179],
[-1.13720713, -0.45453022],
[-1.23602833, -0.74178243],
[-0.38369551, -0.72382916],
[-0.87780149,  0.44313292],
[-1.70542902, -0.31090412],
[-0.6554538 , -0.7328058 ],
[-1.47072867, -0.19420791],
[-0.87780149, -0.83154874],
[-0.77898029, -1.13675421],
[-0.87780149,  0.74833839],
[-1.13720713, -0.23011443],
[-0.49486935, -0.89438516],
[-0.81603824,  0.10202093],
[-1.45837602, -0.55327317],
[-0.6060432 , -0.54429654],
[-0.71721705,  0.19178724],
[-0.92721209, -0.54429654],
[-0.34663756, -0.52634327],
[-0.96427004, -0.93926832],
[-1.71778167, -0.88540853],
[-1.90307141,  1.26000639],
[-0.59369055,  0.08406767],
[-1.53249192,  0.30848345],
[-1.96483466, -1.43298305],
[-1.13720713, -0.849502  ],
[-2.43423535, -0.74178243],
[-1.45837602, -0.77768895],
[-0.71721705, -0.65201611],
[-0.28487431,  0.98173081],
[-1.23602833,  0.98173081],
[-1.91542406,  0.05713777],
[-1.77954492, -0.25704433],
[-0.71721705,  1.87939396],
[ 0.06099988,  3.10919247],
[-1.39661278,  1.77167438],
[-1.14955978, -0.15830138],
[-0.7048644 , -0.72382916],
[-1.49543397, -0.18523128],
[-0.77898029, -0.63406285],
[-1.18661773,  1.76269775],
[-0.17370046, -0.88540853],
[-0.14899516,  0.58675903],
[-0.23546371, -0.02365191],
[-0.37134286,  1.08945039],
```

```
           [-0.6060432 , -0.98415148],
           [-0.49486935,  0.11099756],
           [-0.92721209,  2.13971627],
           [-0.5813379 ,  2.84887015],
           [ 0.60451647,  1.12535692],
           [-0.19840576,  0.55982913],
           [-0.08723191,  0.42517966],
           [ 0.44393202,  0.20076388],
           [ 0.64157442,  0.74833839],
           [ 0.76510091,  2.34617879],
           [-0.92721209,  1.38567923],
           [ 0.19687903,  1.10740365],
           [ 1.0862698 ,  2.42696848],
           [-0.16134781,  2.04097332],
           [ 0.39452142,  0.81117481],
           [ 0.09805783,  1.40363249],
           [ 0.61686912,  0.70345524],
           [-0.26016901,  0.29950682],
           [ 0.13511578, -0.3916938 ],
           [ 0.28334758,  0.8650346 ],
           [-0.51957465, -0.93926832],
           [ 0.20923168,  2.56161795],
           [ 1.0368592 ,  1.60111838],
           [-0.6801591 ,  0.62266555],
           [ 1.65449169, -0.58917969],
           [ 0.59216382, -0.59815632],
           [-0.79133294,  1.34079607],
           [ 0.85156946,  0.82912808],
           [-0.18605311,  0.83810471],
           [-0.05017396,  0.99968408],
           [ 0.96274331,  0.3802965 ],
           [ 0.90098006,  1.81655754],
           [ 0.55510587,  1.22409986],
           [-0.22311106,  0.92787102],
           [ 0.71569031,  0.21871714],
           [ 0.49334262,  2.03199669],
           [-0.98897534,  0.62266555],
           [-0.28487431,  0.04816114],
           [ 1.43214399,  0.15588072],
           [ 0.87627476,  2.974543  ],
           [ 0.49334262,  1.41260912],
           [ 0.33275817,  1.74474449],
           [ 0.20923168,  0.22769377],
           [ 1.39508604,  1.58316512]])
```

[ ]: #after this go through the theory notes to know when to use standardization and
     →normalization