

LAB 2: Red Wine Quality Data Analysis Using Numpy Part- II

Name : ANNAPOORNIMA

ROLL NO : 225229101

IMPORT NECESSARY MODULES

```
In [1]: import numpy as np
```

```
In [3]: wines=np.genfromtxt("winequality-red.csv",delimiter=";",skip_header=1)
```

Numpy Aggergation Methods

Find sum of all residual sugar values

```
In [26]: y=wines[:,3]  
sum(y)
```

```
Out[26]: 4059.5500000000003
```

Find sums of every feature value. Thereare 12 feature altogether

```
In [21]: sum(wines)
```

```
Out[21]: array([13303.1    ,  843.985   ,  433.29    ,  4059.55    ,  139.859   ,  
                25384.    ,  74302.    ,  1593.79794,  5294.47    ,  1052.38    ,  
                16666.35  ,  9012.     ])
```

Find the sum of every row

```
In [25]: z=wines[:,:].sum(axis=1)  
z
```

```
Out[25]: array([ 74.5438 , 123.0548 ,  99.699  , ..., 100.48174, 105.21547,  
                92.49249])
```

What is its size?

```
In [29]: wines.shape[0]
```

```
Out[29]: 1599
```

What is the maximum residual sugar value in red wines data?

```
In [38]: a=wines[:,3]
a=y.astype('int32')
a
```

```
Out[38]: array([1, 2, 2, ..., 2, 2, 3])
```

find its maximum residual sugar value

```
In [39]: max(a)
```

```
Out[39]: 15
```

What is the minimum residual sugar value in red wines data?

```
In [40]: min(a)
```

```
Out[40]: 0
```

What is the average residual sugar value in red wines data?

```
In [42]: np.mean(y)
```

```
Out[42]: 2.53880550343965
```

What is 25 percentile residual sugar value?

```
In [60]: np.percentile(wines[:,3],25)
```

```
Out[60]: 1.9
```

What is 75 percentile residual sugar value?

```
In [61]: np.percentile(wines[:,3],75)
```

```
Out[61]: 2.6
```

Find the average of each feature value

```
In [62]: np.mean(wines,axis=0)
```

```
Out[62]: array([ 8.31963727,  0.52782051,  0.27097561,  2.5388055 ,  0.08746654,  
                15.87492183, 46.46779237,  0.99674668,  3.3111132 ,  0.65814884,  
                10.42298311,  5.63602251])
```

NumPy Array Comparisons

Show all wines with quality > 5

```
In [45]: wines[:,11] > 5
```

```
Out[45]: array([False, False, False, ...,  True, False,  True])
```

Show all wines with quality > 7

```
In [46]: wines[:,11] > 7
```

```
Out[46]: array([False, False, False, ..., False, False, False])
```

check if any wines value is True for the condition quality > 7

```
In [63]: checking=wines[:,11]>7  
True in checking
```

```
Out[63]: True
```

Show first 3 rows where wine quality > 7, call it high_quality

```
In [47]: high_quality = wines[:,11] > 7  
high_quality
```

```
Out[47]: array([False, False, False, ..., False, False, False])
```

Show only top 3 rows and all columns of high_quality wines data

```
In [48]: high_quality = wines[:,11] > 7
         wines[high_quality,:][:3,:]
```

```
Out[48]: array([[7.900e+00, 3.500e-01, 4.600e-01, 3.600e+00, 7.800e-02, 1.500e+01,
                3.700e+01, 9.973e-01, 3.350e+00, 8.600e-01, 1.280e+01, 8.000e+00],
                [1.030e+01, 3.200e-01, 4.500e-01, 6.400e+00, 7.300e-02, 5.000e+00,
                1.300e+01, 9.976e-01, 3.230e+00, 8.200e-01, 1.260e+01, 8.000e+00],
                [5.600e+00, 8.500e-01, 5.000e-02, 1.400e+00, 4.500e-02, 1.200e+01,
                8.800e+01, 9.924e-01, 3.560e+00, 8.200e-01, 1.290e+01, 8.000e+00]])
```

Show wines with a lot of alcohol > 10 and high wine quality > 7

```
In [50]: high_quality_and_alcohol = (wines[:,10] > 10) & (wines[:,11] > 7)
```

show only alcohol and wine quality columns

```
In [51]: high_quality_and_alcohol = (wines[:,10] > 10) & (wines[:,11] > 7)
         wines[high_quality_and_alcohol,10:]
```

```
Out[51]: array([[12.8,  8. ],
                [12.6,  8. ],
                [12.9,  8. ],
                [13.4,  8. ],
                [11.7,  8. ],
                [11. ,  8. ],
                [11. ,  8. ],
                [14. ,  8. ],
                [12.7,  8. ],
                [12.5,  8. ],
                [11.8,  8. ],
                [13.1,  8. ],
                [11.7,  8. ],
                [14. ,  8. ],
                [11.3,  8. ],
                [11.4,  8. ]])
```

Combining NumPy Arrays

Combine red wine and white wine data

Open white wine dataset

```
In [56]: white_wines = np.genfromtxt("winequality-white .csv", delimiter=";", skip_header=
```

Show size of white_wines

```
In [57]: white_wines.shape
```

```
Out[57]: (4898, 12)
```

combine wines and white_wines data frames using vstack and call it all_wines

```
In [58]: all_wines = np.vstack((wines, white_wines))  
all_wines.shape
```

```
Out[58]: (6497, 12)
```

Combine wines and white_wines data frames using concatenate method

```
In [59]: np.concatenate((wines, white_wines), axis=0)
```

```
Out[59]: array([[ 7.4 ,  0.7 ,  0.   , ...,  0.56,  9.4 ,  5.   ],  
                [ 7.8 ,  0.88,  0.   , ...,  0.68,  9.8 ,  5.   ],  
                [ 7.8 ,  0.76,  0.04, ...,  0.65,  9.8 ,  5.   ],  
                ...,  
                [ 6.5 ,  0.24,  0.19, ...,  0.46,  9.4 ,  6.   ],  
                [ 5.5 ,  0.29,  0.3 , ...,  0.38, 12.8 ,  7.   ],  
                [ 6.   ,  0.21,  0.38, ...,  0.32, 11.8 ,  6.   ]])
```

Matrix Operations and Reshape

Find Transpose of wines and print its size

```
In [64]: tran=wines.T  
tran.shape
```

```
Out[64]: (12, 1599)
```

Convert wines data into 1D array

```
In [65]: wines.ravel()
```

```
Out[65]: array([ 7.4 ,  0.7 ,  0.   , ...,  0.66, 11.   ,  6.   ])
```

show size

```
In [66]: wines.ravel().shape
```

```
Out[66]: (19188,)
```

Reshape second row of wines into a 2-dimensional array with 2 rows and 6 columns

```
In [67]: wines[1].reshape((2,6))
```

```
Out[67]: array([[ 7.8   ,  0.88   ,  0.    ,  2.6   ,  0.098 , 25.    ],
                [67.    ,  0.9968,  3.2   ,  0.68   ,  9.8   ,  5.    ]])
```

Sort alcohol column Ascending Order

```
In [69]: sorted_alcohol=np.sort(wines[:,-2])
```

```
In [71]: sorted_alcohol
```

```
Out[71]: array([ 8.4,  8.4,  8.5, ..., 14. , 14. , 14.9])
```

Make sorting to take place in-place

```
In [72]: wines[:,-2].sort()
```

Show top 10 rows

```
In [73]: wines[:,-2]
```

```
Out[73]: array([ 8.4,  8.4,  8.5, ..., 14. , 14. , 14.9])
```

Sort alcohol column Descending Order

```
In [74]: sorted_alcohol_desc=np.sort(wines[:,-2])[::-1]
```

```
In [75]: sorted_alcohol_desc
```

```
Out[75]: array([14.9, 14. , 14. , ...,  8.5,  8.4,  8.4])
```

Will original data be modified?. Check top 10 rows

```
In [76]: wines[:,-2]
```

```
Out[76]: array([ 8.4,  8.4,  8.5, ..., 14. , 14. , 14.9])
```

```
In [ ]:
```

