

Lab_WineData

May 5, 2020

0.1 Red Wine Quality Data Analysis using NumPy Part-I

```
In [ ]: '''  
        Wine quality dataset 11 input features and 1 output feature  
  
        1 - fixed acidity  
        2 - volatile acidity  
        3 - citric acid  
        4 - residual sugar  
        5 - chlorides  
        6 - free sulfur dioxide  
        7 - total sulfur dioxide  
        8 - density  
        9 - pH  
        10 - sulphates  
        11 - alcohol  
        Output variable (based on sensory data):  
        12 - quality (score between 0 and 10)'''
```

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

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

What is its shape

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

How many wine data rows here?

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

How many wine data columns here?

```
In [ ]: wines.shape[1]
```

How many dimensions?

```
In [ ]: wines.ndim
```

What is the type of data?

```
In [ ]: type(wines)
```

What is the data type of wines data?

```
In [ ]: wines.dtype
```

Show top 5 rows

```
In [ ]: #wines[:5, :]
```

What is the value at 3rd row, 4th column of wine data?

```
In [ ]: wines[2,3]
```

Select first 3 items in 4th column

```
In [ ]: wines[:3, 3]
```

Show 1st column

```
In [ ]: wines[:, 0]
```

Show 2nd row

```
In [ ]: wines[1, :]
```

Select items from rows 1 to 3 and 5th column

```
In [ ]: wines[1:4, 4]
```

Select entire array

```
In [ ]: wines[:, :]
```

Change 1st value in wines to 100

```
In [ ]: wines[0,0]
```

```
In [ ]: wines[0,0] = 100
```

```
In [ ]: wines[0,0]
```

```
In [ ]: # change it back
        wines[0,0] = 7.4
```

0.1.1 1-Dimensional Numpy Arrays

Select 4th row all column values

```
In [ ]: third_wine = wines[3, :]
```

```
In [ ]: third_wine
```

```
In [ ]: third_wine[1]
```

Convert one datatype to another

```
In [ ]: #convert to int  
        wines.astype(int)
```

0.1.2 Vectorization Operations

Increase wine quality score (output variable) by 10

```
In [ ]: # check values first  
        wines[:, 11]
```

```
In [ ]: wines[:, 11] += 10
```

```
In [ ]: wines[:, 11]
```

Multiply alcohol of all wine data by 3 times

```
In [ ]: wines[:, 10] *= 3
```

```
In [ ]: wines[:, 10]
```

Add quality column by itself

```
In [ ]: # It will produce a new array  
        wines[:, 11] + wines[:, 11]
```

Multiply alcohol and wine quality columns. It will perform element wise multiplication

```
In [ ]: wines[:,10] * wines[:,11]
```

0.1.3 Broadcasting

Add every row of wines data with a random array of values

```
In [ ]: rand_array = np.random.rand(12) #here 12, becos there are 12 columns
```

```
In [ ]: rand_array
```

```
In [ ]: wines + rand_array
```

```
In [ ]:
```