

## ▼ Chapter 5 - Basic Math and Statistics

### Segment 7 - Transforming dataset distributions

```
import numpy as np
import pandas as pd
import scipy

import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sb

import sklearn
from sklearn import preprocessing
from sklearn.preprocessing import scale

%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')
```

## ▼ Normalizing and transforming features with MinMaxScalar() and fit\_transform()

```
address = 'C:/Users/Lillian/Desktop/ExerciseFiles/Data/mtcars.csv'

cars = pd.read_csv(address)
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']

mpg = cars.mpg
plt.plot(mpg)
```

```
[<matplotlib.lines.Line2D at 0x1b2b09f1128>]
```



```
cars[['mpg']].describe()
```

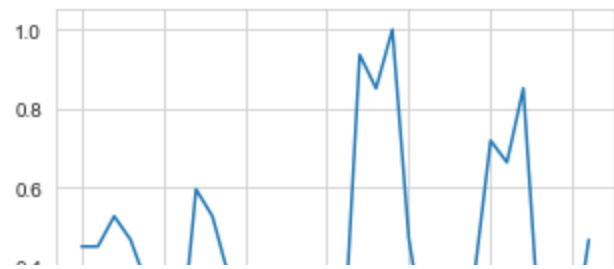
	mpg
<b>count</b>	32.000000
<b>mean</b>	20.090625
<b>std</b>	6.026948
<b>min</b>	10.400000
<b>25%</b>	15.425000
<b>50%</b>	19.200000
<b>75%</b>	22.800000
<b>max</b>	33.900000

```
mpg_matrix = mpg.values.reshape(-1,1)
```

```
scaled = preprocessing.MinMaxScaler()
```

```
scaled_mpg = scaled.fit_transform(mpg_matrix)  
plt.plot(scaled_mpg)
```

```
[<matplotlib.lines.Line2D at 0x1b2b0ba5f28>]
```

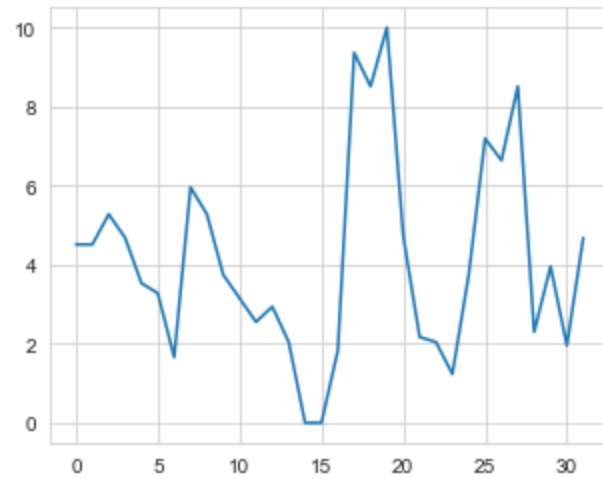


```
scaled = preprocessing.MinMaxScaler(feature_range=(0,10))
```

```
scaled_mpg = scaled.fit_transform(mpg_matrix)
```

```
plt.plot(scaled_mpg)
```

```
[<matplotlib.lines.Line2D at 0x1b2b0c0ac18>]
```

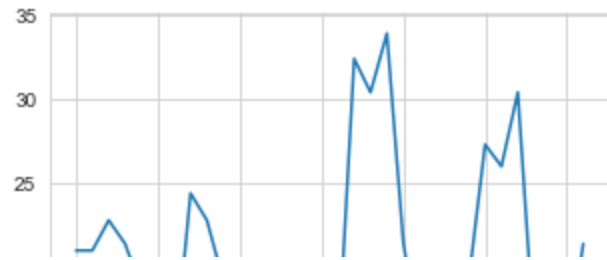


## ▼ Using scale() to scale your features

```
standardized_mpg = scale(mpg, axis=0, with_mean=False, with_std=False)
```

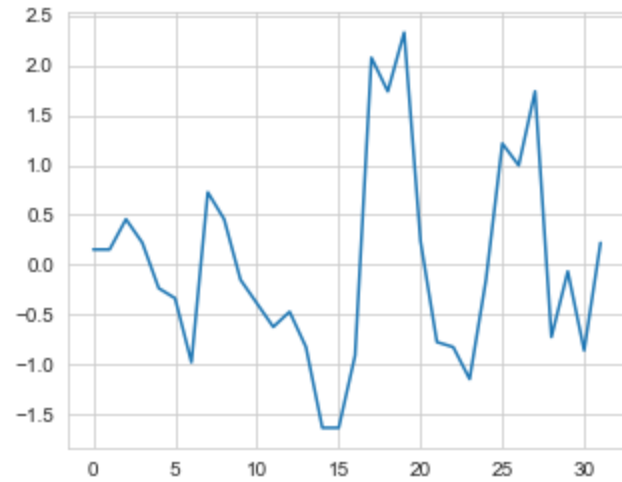
```
plt.plot(standardized_mpg)
```

```
[<matplotlib.lines.Line2D at 0x1b2b0c6e320>]
```



```
standardized_mpg = scale(mpg)
plt.plot(standardized_mpg)
```

```
[<matplotlib.lines.Line2D at 0x1b2b0b0a978>]
```



<http://goo.gl/tuEWkD>

[Colab paid products](#) - [Cancel contracts here](#)

---

