

Feature Scaling

March 19, 2023

1 Feature Scaling

1.1 Standardization

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

```
[2]: df=sns.load_dataset("tips")
```

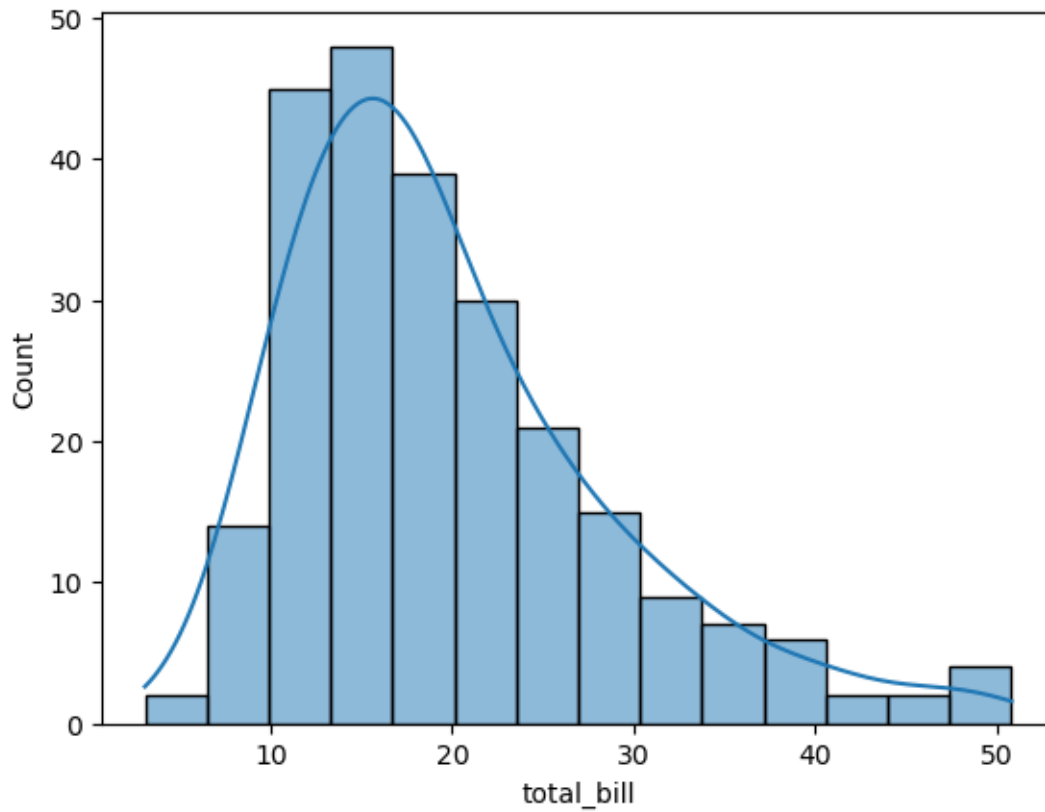
```
[8]: df.head()
```

```
[8]:
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

```
[10]: sns.histplot(df.total_bill,kde=True)
```

```
[10]: <AxesSubplot: xlabel='total_bill', ylabel='Count'>
```



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

```
[5]: scale=StandardScaler()
```

```
[11]: ans=pd.DataFrame(scale.  
    ↪fit_transform(df[["total_bill","tip","size"]]),columns=["total_bill","tip","size"])
```

```
[12]: ans
```

```
[12]:
```

	total_bill	tip	size
0	-0.314711	-1.439947	-0.600193
1	-1.063235	-0.969205	0.453383
2	0.137780	0.363356	0.453383
3	0.438315	0.225754	-0.600193
4	0.540745	0.443020	1.506958
..
239	1.040511	2.115963	0.453383
240	0.832275	-0.722971	-0.600193
241	0.324630	-0.722971	-0.600193
242	-0.221287	-0.904026	-0.600193
243	-0.113229	0.001247	-0.600193

[244 rows x 3 columns]

1.2 Normalization

```
[19]: df
```

```
[19]:
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
..
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

[244 rows x 7 columns]

```
[20]: from sklearn.preprocessing import MinMaxScaler
```

```
[21]: scale=MinMaxScaler()
```

```
[22]: ans=pd.DataFrame(scale.  
    ↪fit_transform(df[["total_bill","tip","size"]]),columns=["total_bill","tip","size"])
```

```
[23]: ans
```

```
[23]:
```

	total_bill	tip	size
0	0.291579	0.001111	0.2
1	0.152283	0.073333	0.4
2	0.375786	0.277778	0.4
3	0.431713	0.256667	0.2
4	0.450775	0.290000	0.6
..
239	0.543779	0.546667	0.4
240	0.505027	0.111111	0.2
241	0.410557	0.111111	0.2
242	0.308965	0.083333	0.2
243	0.329074	0.222222	0.2

[244 rows x 3 columns]

```
[24]: df
```

```
[24]:
```

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
..
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

[244 rows x 7 columns]

1.3 Unit Vector

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

```
[27]: pd.
↳ DataFrame(normalize(df[["total_bill", "tip", "size"]]), columns=["total_bill", "tip", "size"])
```

```
[27]:
```

	total_bill	tip	size
0	0.991416	0.058936	0.116706
1	0.949178	0.152383	0.275390
2	0.976766	0.162717	0.139472
3	0.986925	0.137953	0.083355
4	0.976825	0.143405	0.158898
..
239	0.974849	0.198798	0.100742
240	0.994629	0.073188	0.073188
241	0.992307	0.087544	0.087544
242	0.989062	0.097130	0.111006
243	0.982064	0.156879	0.104586

[244 rows x 3 columns]

```
[ ]:
```