

```
In [ ]: import pandas as pd
```

Q3 Load the data from wine dataset. Check whether all attributes are standardized or not (mean is 0 and standard deviation is 1). If not, standardize the attributes. Do the same with Iris dataset.

```
In [ ]: data=pd.read_csv('winequalityN.csv')
# Drop categorical type column
data.drop(columns=['type'],inplace=True)
data2=data.copy()
display(data)
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7.0	0.270	0.36	20.7	0.045	45.0	170.0	1.00100	3.00	0.45	
1	6.3	0.300	0.34	1.6	0.049	14.0	132.0	0.99400	3.30	0.49	
2	8.1	0.280	0.40	6.9	0.050	30.0	97.0	0.99510	3.26	0.44	1
3	7.2	0.230	0.32	8.5	0.058	47.0	186.0	0.99560	3.19	0.40	
4	7.2	0.230	0.32	8.5	0.058	47.0	186.0	0.99560	3.19	0.40	
...	
6492	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	1
6493	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	NaN	1
6494	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	1
6495	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	1
6496	6.0	0.310	0.47	3.6	0.067	18.0	42.0	0.99549	3.39	0.66	1

6497 rows × 12 columns

```
In [ ]: # check if attributes are standardized or not
d=data.describe()
for i in d.columns:
    if d[i]['mean']!=0 or d[i]['std']!=1 :
        print(i, ' column is not standardized.')
```

```
fixed acidity  column is not standarized.
volatile acidity  column is not standarized.
citric acid  column is not standarized.
residual sugar  column is not standarized.
chlorides  column is not standarized.
free sulfur dioxide  column is not standarized.
total sulfur dioxide  column is not standarized.
density  column is not standarized.
pH  column is not standarized.
sulphates  column is not standarized.
alcohol  column is not standarized.
quality  column is not standarized.
```

In []:

```
# second method
from sklearn.preprocessing import StandardScaler
ss = StandardScaler()
data=ss.fit_transform(data2)
data=pd.DataFrame(data,columns=data2.columns)
display(data)
display(data.describe())
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	
0	-0.167030	-0.423303	0.284180	3.206483	-0.315173	0.815565	0.959976	2.102214	-1.356
1	-0.706883	-0.241083	0.146489	-0.808012	-0.200996	-0.931107	0.287618	-0.232332	0.507
2	0.681310	-0.362563	0.559560	0.305958	-0.172452	-0.029599	-0.331660	0.134525	0.258
3	-0.012786	-0.666262	0.008799	0.642251	0.055902	0.928254	1.243074	0.301278	-0.176
4	-0.012786	-0.666262	0.008799	0.642251	0.055902	0.928254	1.243074	0.301278	-0.176
...
6492	-0.784004	1.581115	-1.643483	-0.723939	0.969318	0.083090	-1.269422	0.067824	1.440
6493	-1.015370	1.277415	-1.505793	-0.681902	0.170079	0.477500	-1.145567	0.141195	1.876
6494	-0.706883	1.034456	-1.299258	-0.660884	0.569698	-0.085943	-1.340197	0.347969	1.254
6495	-1.015370	1.854445	-1.368103	-0.723939	0.541154	0.083090	-1.269422	0.257923	2.187
6496	-0.938248	-0.180343	1.041476	-0.387646	0.312800	-0.705730	-1.304809	0.264593	1.067

6497 rows × 12 columns

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	
count	6.487000e+03	6489.000000	6.494000e+03	6.495000e+03	6.495000e+03	6.497000e+03	6
mean	4.030827e-16	0.000000	4.726739e-16	-1.400300e-16	-7.001499e-17	-8.749179e-17	
std	1.000077e+00	1.000077	1.000077e+00	1.000077e+00	1.000077e+00	1.000077e+00	
min	-2.634928e+00	-1.577361	-2.194244e+00	-1.018195e+00	-1.342766e+00	-1.663583e+00	
25%	-6.297608e-01	-0.666262	-4.731166e-01	-7.659755e-01	-5.149831e-01	-7.620742e-01	
50%	-1.670299e-01	-0.301823	-6.004599e-02	-5.137559e-01	-2.580848e-01	-8.594301e-02	
75%	3.728228e-01	0.366316	4.907148e-01	5.581775e-01	2.557116e-01	5.901882e-01	
max	6.696812e+00	7.533628	9.234043e+00	1.268574e+01	1.584087e+01	1.456357e+01	

In []:

```
import seaborn as sns
iris=sns.load_dataset('iris')
display(iris)
display(iris.describe())
iris.drop(columns=['species'],inplace=True)
data=ss.fit_transform(iris)
data=pd.DataFrame(data,columns=iris.columns)
display(data)
display(data.describe())
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa
...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

	sepal_length	sepal_width	petal_length	petal_width
0	-0.900681	1.019004	-1.340227	-1.315444
1	-1.143017	-0.131979	-1.340227	-1.315444
2	-1.385353	0.328414	-1.397064	-1.315444
3	-1.506521	0.098217	-1.283389	-1.315444
4	-1.021849	1.249201	-1.340227	-1.315444

	sepal_length	sepal_width	petal_length	petal_width
...
145	1.038005	-0.131979	0.819596	1.448832
146	0.553333	-1.282963	0.705921	0.922303
147	0.795669	-0.131979	0.819596	1.053935
148	0.432165	0.788808	0.933271	1.448832
149	0.068662	-0.131979	0.762758	0.790671
	sepal_length	sepal_width	petal_length	petal_width
count	1.500000e+02	1.500000e+02	1.500000e+02	1.500000e+02
mean	-4.736952e-16	-7.815970e-16	-4.263256e-16	-4.736952e-16
std	1.003350e+00	1.003350e+00	1.003350e+00	1.003350e+00
min	-1.870024e+00	-2.433947e+00	-1.567576e+00	-1.447076e+00
25%	-9.006812e-01	-5.923730e-01	-1.226552e+00	-1.183812e+00
50%	-5.250608e-02	-1.319795e-01	3.364776e-01	1.325097e-01
75%	6.745011e-01	5.586108e-01	7.627583e-01	7.906707e-01
max	2.492019e+00	3.090775e+00	1.785832e+00	1.712096e+00

In []: