Introduction to Machine Learning Assignment

Problem Statement: You work in XYZ Company as a Python. The company officials want you to build a data science model. Tasks to be performed:

1. Using sklearn import the wine dataset
2. Split the data into train and test set
3. Train the model
4. Make Predictions
5. Check the performance of the model using r2_score.

```python
In [2]:  import os
         os.chdir("C:\\Users\\veena\\OneDrive\\Desktop\\intellipaat assighnment pdf s")
```

```python
In [19]:  import pandas as pd
          import matplotlib.pyplot as plt
          %matplotlib inline
          df=pd.read_csv('wine.csv')
          df.head()
```

Out[19]:

	Wine	Alcohol	Malic.acid	Ash	Acl	Mg	Phenols	Flavanoids	Nonflavanoid.phenols	Proanth
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82

```python
In [30]:  #wine.info()
          df.corr()
```

Out[30]:

	Wine	Alcohol	Malic.acid	Ash	Acl	Mg	Phenols
Wine	1.000000	-0.328222	0.437776	-0.049643	0.517859	-0.209179	-0.719163
Alcohol	-0.328222	1.000000	0.094397	0.211545	-0.310235	0.270798	0.289101
Malic.acid	0.437776	0.094397	1.000000	0.164045	0.288500	-0.054575	-0.335167
Ash	-0.049643	0.211545	0.164045	1.000000	0.443367	0.286587	0.128980
Acl	0.517859	-0.310235	0.288500	0.443367	1.000000	-0.083333	-0.321113
Mg	-0.209179	0.270798	-0.054575	0.286587	-0.083333	1.000000	0.214401
Phenols	-0.719163	0.289101	-0.335167	0.128980	-0.321113	0.214401	1.000000
Flavanoids	-0.847498	0.236815	-0.411007	0.115077	-0.351370	0.195784	0.864564
Nonflavanoid.phenols	0.489109	-0.155929	0.292977	0.186230	0.361922	-0.256294	-0.449935
Proanth	-0.499130	0.136698	-0.220746	0.009652	-0.197327	0.236441	0.612413
Color.int	0.265668	0.546364	0.248985	0.258887	0.018732	0.199950	-0.055136
Hue	-0.617369	-0.071747	-0.561296	-0.074667	-0.273955	0.055398	0.433681
OD	-0.788230	0.072343	-0.368710	0.003911	-0.276769	0.066004	0.699949
Proline	-0.633717	0.643720	-0.192011	0.223626	-0.440597	0.393351	0.498115

In [40]:
```
X= df[['Alcohol','Malic.acid','Phenols']]
y= df['Wine']
```

In [37]: `X`

Out[37]:

	Alcohol	Malic.acid	Phenols
0	14.23	1.71	2.80
1	13.20	1.78	2.65
2	13.16	2.36	2.80
3	14.37	1.95	3.85
4	13.24	2.59	2.80
...	...	...	...
173	13.71	5.65	1.68
174	13.40	3.91	1.80
175	13.27	4.28	1.59
176	13.17	2.59	1.65
177	14.13	4.10	2.05

178 rows × 3 columns

In [41]: `y`

```
Out[41]: 0      1
         1      1
         2      1
         3      1
         4      1
                ..
         173    3
         174    3
         175    3
         176    3
         177    3
         Name: Wine, Length: 178, dtype: int64
```

In [9]:
```python
from sklearn.model_selection import train_test_split
```

In [42]:
```python
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2)
```

In [43]:
```python
len(X_train)
```

Out[43]: 142

In [44]:
```python
len(X_test)
```

Out[44]: 36

In [45]:
```python
from sklearn.linear_model import LinearRegression
clf=LinearRegression()
```

In [46]:
```python
clf.fit(X_train,y_train)
```

Out[46]:
```
LinearRegression()
```

In [47]:
```python
clf.predict(X_test)
```

Out[47]:
```
array([0.84259044, 1.62767085, 1.22640997, 2.50543042, 2.62297679,
       2.62983175, 0.87308151, 1.20057493, 1.50710827, 0.90729738,
       2.65800264, 1.53593931, 2.01958089, 2.9457584 , 2.77738542,
       2.3144795 , 2.7836144 , 2.4960967 , 1.97369406, 2.0214592 ,
       1.76197157, 1.81838971, 2.95911629, 2.20495359, 1.13782083,
       2.39493733, 1.88617943, 2.32020522, 2.4378826 , 1.69554665,
       1.51188673, 1.68169212, 1.90210912, 1.17596398, 1.79570614,
       1.94971565])
```

In [50]:
```python
clf.score(X_test,y_test)
```

Out[50]: 0.5638133121649735

In [ ]: