**Case Study using Wine data set**



Certification and quality assessment are important issues within the wine industry. Since, the wine quality is assessed by the physio-chemical and human sensory tests, data are collected to perform a wine quality assessment. Regression approach is usually plied to predict the quality of the wine.

|  |
| --- |
|  |
| **Details of the wine data set** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Set Characteristics:** | Multivariate | **Number of Instances:** | 179 | **Associated Tasks:** | Classification |
| **Attribute Characteristics:** | Real | **Number of Attributes:** | 13 | **Date Donated** | NA |

Attribute Information:

1. Malic acid
2. Ash
3. Alcalinity of ash
4. Magnesium
5. Total phenol
6. Flavanoids
7. Nonflavanoid phenols
8. Proanthocyanins
9. Color intensity
10. Hue
11. OD280/OD315 of diluted wines
12. Proline

class: 3

Load the “Wine dataset.xls” and perform the following questions:

1. Describe the data for its statistical properties and discuss what you infer from it.
2. Find out the distribution of the given data set.
3. Conduct an experiment to find out whether the sampled distribution derived from the 3 classes are well separated or not.
4. Design and implement a regression model to predict the quality of the wine. Comment on which feature would result in the accurate prediction.