Homework 05

The wine dataset is a multi-class classification dataset which contains three different wine categories and 13 continuous-valued features, for a total of 178 observations.

The goal is to classify an unlabeled wine according to its characteristic features.

In [2]:
```
from sklearn import datasets
dataset = datasets.load_wine()

X = dataset.data
y = dataset.target

print("\nX.shape =",  X.shape)
print("\ny.shape =",  y.shape)
print("\nwine categories:\n", dataset['target'])
print("\nfeatures names:\n", dataset['feature_names'])
```

```
X.shape = (178, 13)

y.shape = (178,)

wine categories:
 [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2]

features names:
 ['alcohol', 'malic_acid', 'ash', 'alcalinity_of_ash', 'magnesium', 'total_phenols', 'flavanoids
', 'nonflavanoid_phenols', 'proanthocyanins', 'color_intensity', 'hue', 'od280/od315_of_diluted_w
ines', 'proline']
```

1. Perform a train-test split on the data using sklearn `train_test_split` with `test_size=0.3` . Name your variables `X_train` , `X_test` , `y_train` , `y_test` . Make sure that your training set contains samples from all the categories.
2. Fit sklearn `LogisticRegression` model to the training data `X_train` , `y_train` , predict the classification labels on the test data `X_test` and use sklearn `classification_report` to evaluate your model against the actual labels `y_test` .
3. Repeat step 2. using sklearn Naive Bayes classifier `GaussianNB` .