In [ ]:

```
# Name : Meet Trivedi
# Student Id; N01520331
```

Lab 9- K Mean Clustering

Import Libraries

In [1]:

```
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt

import sklearn
from sklearn.preprocessing import scale
import sklearn.metrics as sm
from sklearn.metrics import confusion_matrix, classification_report
from sklearn.cluster import KMeans
from mpl_toolkits.mplot3d import Axes3D
from sklearn import datasets
%matplotlib inline
```

Get the Data(Wine dataset)

In [12]:

```
df =  datasets.load_wine()
df
```

Out[12]:


```
{'data': array([[1.423e+01, 1.710e+00, 2.430e+00, ..., 1.040e+00, 3.920e+00,
         1.065e+03],
        [1.320e+01, 1.780e+00, 2.140e+00, ..., 1.050e+00, 3.400e+00,
         1.050e+03],
        [1.316e+01, 2.360e+00, 2.670e+00, ..., 1.030e+00, 3.170e+00,
         1.185e+03],
        ...,
        [1.327e+01, 4.280e+00, 2.260e+00, ..., 5.900e-01, 1.560e+00,
         8.350e+02],
        [1.317e+01, 2.590e+00, 2.370e+00, ..., 6.000e-01, 1.620e+00,
         8.400e+02],
        [1.413e+01, 4.100e+00, 2.740e+00, ..., 6.100e-01, 1.600e+00,
         5.600e+02]]),
 'target': array([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, 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, 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]),
 'frame': None,
 'target_names': array(['class_0', 'class_1', 'class_2'], dtype='<U7'),
 'DESCR': '.. _wine_dataset:\n\nWine recognition dataset\n------------------
------\n\n**Data Set Characteristics:**\n\n    :Number of Instances: 178 (50
in each of three classes)\n    :Number of Attributes: 13 numeric, predictive
attributes and the class\n    :Attribute Information:\n \t\t- Alcohol\n \t\t
- Malic acid\n \t\t- Ash\n\t\t- Alcalinity of ash  \n \t\t- Magnesium\n\t\t-
Total phenols\n \t\t- Flavanoids\n \t\t- Nonflavanoid phenols\n \t\t- Proant
hocyanins\n\t\t- Color intensity\n \t\t- Hue\n \t\t- OD280/OD315 of diluted
wines\n \t\t- Proline\n\n    - class:\n            - class_0\n            -
class_1\n            - class_2\n\t\t\n    :Summary Statistics:\n    \n    ==
========================== ==== ===== ======= =====\n
Min   Max   Mean     SD\n    ========================== ==== ===== ======
= =====\n    Alcohol:                    11.0  14.8    13.0   0.8\n    Mal
ic Acid:                  0.74  5.80    2.34  1.12\n    Ash:
1.36  3.23    2.36  0.27\n    Alcalinity of Ash:           10.6  30.0    1
9.5   3.3\n    Magnesium:                  70.0 162.0    99.7  14.3\n    T
otal Phenols:               0.98  3.88    2.29  0.63\n    Flavanoids:
0.34  5.08    2.03  1.00\n    Nonflavanoid Phenols:         0.13  0.66    0.
36  0.12\n    Proanthocyanins:             0.41  3.58    1.59  0.57\n    Co
lour Intensity:             1.3  13.0     5.1   2.3\n    Hue:
0.48  1.71    0.96  0.23\n    OD280/OD315 of diluted wines: 1.27  4.00    2.
61  0.71\n    Proline:                     278  1680     746   315\n    ==
========================== ==== ===== ======= =====\n\n    :Missing Attribu
te Values: None\n    :Class Distribution: class_0 (59), class_1 (71), class_
2 (48)\n    :Creator: R.A. Fisher\n    :Donor: Michael Marshall (MARSHALL%PL
U@io.arc.nasa.gov)\n    :Date: July, 1988\n\nThis is a copy of UCI ML Wine r
ecognition datasets.\nhttps://archive.ics.uci.edu/ml/machine-learning-databa
ses/wine/wine.data\n\nThe data is the results of a chemical analysis of wine
s grown in the same\nregion in Italy by three different cultivators. There a
```

```
re thirteen different\nmeasurements taken for different constituents found i
n the three types of\nwine.\n\nOriginal Owners: \n\nForina, M. et al, PARVUS
- \nAn Extendible Package for Data Exploration, Classification and Correlati
on. \nInstitute of Pharmaceutical and Food Analysis and Technologies,\nVia B
rigata Salerno, 16147 Genoa, Italy.\n\nCitation:\n\nLichman, M. (2013). UCI
Machine Learning Repository\n[https://archive.ics.uci.edu/ml]. Irvine, CA: U
niversity of California,\nSchool of Information and Computer Science. \n\n..
topic:: References\n\n  (1) S. Aeberhard, D. Coomans and O. de Vel, \n  Comp
arison of Classifiers in High Dimensional Settings, \n  Tech. Rep. no. 92-0
2, (1992), Dept. of Computer Science and Dept. of  \n  Mathematics and Stati
stics, James Cook University of North Queensland. \n  (Also submitted to Tec
hnometrics). \n\n  The data was used with many others for comparing various
\n  classifiers. The classes are separable, though only RDA \n  has achieved
100% correct classification. \n  (RDA : 100%, QDA 99.4%, LDA 98.9%, 1NN 96.
1% (z-transformed data)) \n  (All results using the leave-one-out technique)
\n\n  (2) S. Aeberhard, D. Coomans and O. de Vel, \n  "THE CLASSIFICATION PE
RFORMANCE OF RDA" \n  Tech. Rep. no. 92-01, (1992), Dept. of Computer Scienc
e and Dept. of \n  Mathematics and Statistics, James Cook University of Nort
h Queensland. \n  (Also submitted to Journal of Chemometrics).\n',
 'feature_names': ['alcohol',
  'malic_acid',
  'ash',
  'alcalinity_of_ash',
  'magnesium',
  'total_phenols',
  'flavanoids',
  'nonflavanoid_phenols',
  'proanthocyanins',
  'color_intensity',
  'hue',
  'od280/od315_of_diluted_wines',
  'proline']}
```

In [19]:

```
v_names = df.feature_names
v_names
```

Out[19]:

```
['alcohol',
 'malic_acid',
 'ash',
 'alcalinity_of_ash',
 'magnesium',
 'total_phenols',
 'flavanoids',
 'nonflavanoid_phenols',
 'proanthocyanins',
 'color_intensity',
 'hue',
 'od280/od315_of_diluted_wines',
 'proline']
```

In [20]:

```
x = scale(df.data)
x
```

Out[20]:

```
array([[ 1.51861254, -0.5622498 ,  0.23205254, ...,  0.36217728,
         1.84791957,  1.01300893],
       [ 0.24628963, -0.49941338, -0.82799632, ...,  0.40605066,
         1.1134493 ,  0.96524152],
       [ 0.19687903,  0.02123125,  1.10933436, ...,  0.31830389,
         0.78858745,  1.39514818],
       ...,
       [ 0.33275817,  1.74474449, -0.38935541, ..., -1.61212515,
        -1.48544548,  0.28057537],
       [ 0.20923168,  0.22769377,  0.01273209, ..., -1.56825176,
        -1.40069891,  0.29649784],
       [ 1.39508604,  1.58316512,  1.36520822, ..., -1.52437837,
        -1.42894777, -0.59516041]])
```

In [21]:

```
y = pd.DataFrame(df.target)
y
```

Out[21]:

	0
0	0
1	0
2	0
3	0
4	0
...	...
173	2
174	2
175	2
176	2
177	2

178 rows × 1 columns

Apply Kmean Clustering

In [34]:

```
clustering = KMeans(n_clusters=3, random_state=5)

clustering.fit(x)
```

Out[34]:

```
KMeans(n_clusters=3, random_state=5)
```

In [35]:

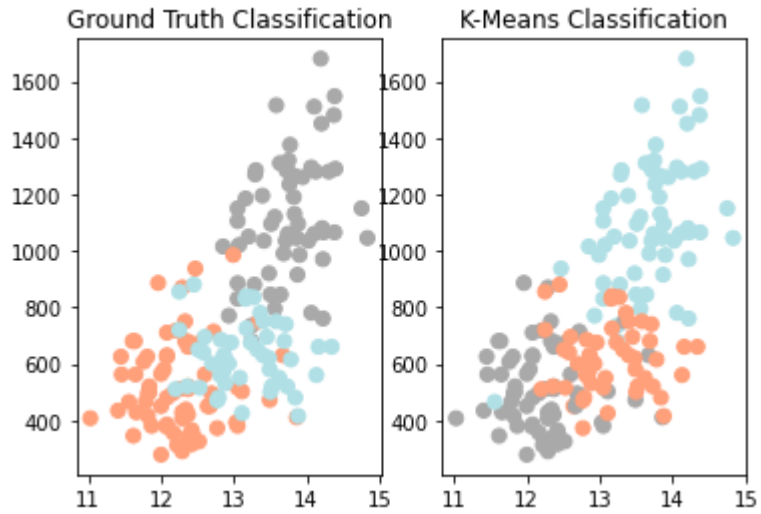
```
wine_df = pd.DataFrame(df.data)
wine_df.columns = ['alcohol','malic_acid', 'ash', 'alcalinity_of_ash', 'magnesium', 'total_
y.columns = ['Targets']
```

In [36]:

```
color_theme = np.array(['darkgray', 'lightsalmon', 'powderblue', 'darkblue', 'red', 'orange

plt.subplot(1,2,1)

plt.scatter(x=wine_df.alcohol, y=wine_df.proline, c=color_theme[df.target], s=50)
plt.title('Ground Truth Classification')

plt.subplot(1,2,2)

plt.scatter(x=wine_df.alcohol, y=wine_df.proline, c=color_theme[clustering.labels_], s=50)
plt.title('K-Means Classification')
```

Out[36]:

```
Text(0.5, 1.0, 'K-Means Classification')
```



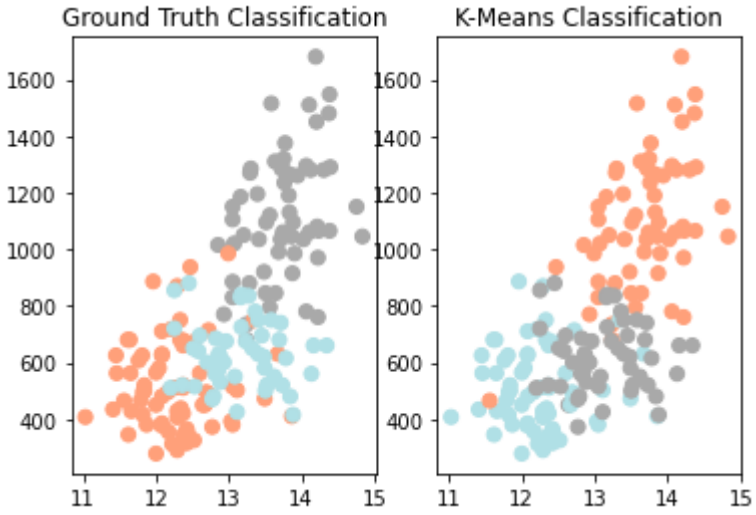
Predictions and Evaluations

In [38]:

```
relabel = np.choose(clustering.labels_, [2, 0, 1 ]).astype(np.int64)

plt.subplot(1,2,1)

plt.scatter(x=wine_df.alcohol, y=wine_df.proline, c=color_theme[df.target], s=50)
plt.title('Ground Truth Classification')

plt.subplot(1,2,2)

plt.scatter(x=wine_df.alcohol, y=wine_df.proline, c=color_theme[relabel], s=50)
plt.title('K-Means Classification')
```

Out[38]:

```
Text(0.5, 1.0, 'K-Means Classification')
```



In [39]:

```
print(classification_report(y, relabel))
```

```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        59
           1       0.05      0.04      0.05        71
           2       0.00      0.00      0.00        48

    accuracy                           0.02       178
   macro avg       0.02      0.01      0.02       178
weighted avg       0.02      0.02      0.02       178
```

In [ ]: