```
from google.colab import drive
import pandas as pd

# Mount Google Drive to access files
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True)
```

```
# Replace the path below with the path to your CSV file in Google Drive
file_path = '/content/drive/MyDrive/00 Machine Learning Notes/Practical lab 9P2/Ex 5 EM Algorithm/WineQT.csv'
```

```
# Load the dataset
df = pd.read_csv(file_path)
display(df.head())
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	Id
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	0
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5	1
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5	2
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6	3

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.mixture import GaussianMixture
```

```
# Step 2: Select relevant features for GMM clustering
# You can customize this list based on your analysis needs
features = ['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',
            'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density',
            'pH', 'sulphates', 'alcohol']
```

```
X = df[features]
X
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65	9.8
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58	9.8
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4
...	...	...	...	...	...	...	...	...	...	...	...
1138	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0
1139	6.8	0.620	0.08	1.9	0.068	28.0	38.0	0.99651	3.42	0.82	9.5
1140	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5
1141	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2
1142	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2

1143 rows × 11 columns

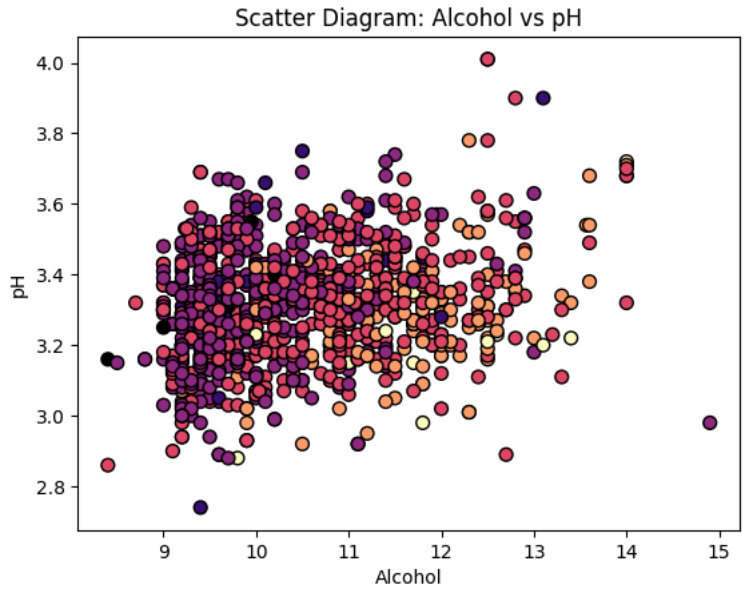
```
y = df.quality  # True labels
y
```

	quality
0	5
1	5
2	5
3	6
4	5
...	...
1138	6
1139	6
1140	5
1141	6
1142	5

1143 rows × 1 columns

dtype: int64

```
# Step 2: Visualize clusters in 2D using two features ('alcohol' and 'pH')
plt.scatter(df['alcohol'], df['pH'], c=y, cmap='magma', edgecolor='k', s=50)  #viridis = colormap
#You can try other colormaps like 'plasma', 'inferno', 'magma', or 'cividis' depending on your preference.
plt.xlabel('Alcohol')
plt.ylabel('pH')
plt.title('Scatter Diagram: Alcohol vs pH')
plt.show()
```




```
# Step 3: Fit Gaussian Mixture Model
n_components = 3  # Choose number of clusters; you can experiment with this value
gmm = GaussianMixture(n_components=n_components, random_state=42)
gmm.fit(X)
```

```
GaussianMixture(n_components=3, random_state=42)
```

```
# Step 4: Cluster assignment
cluster_labels = gmm.predict(X)
df['Cluster'] = cluster_labels
df
```

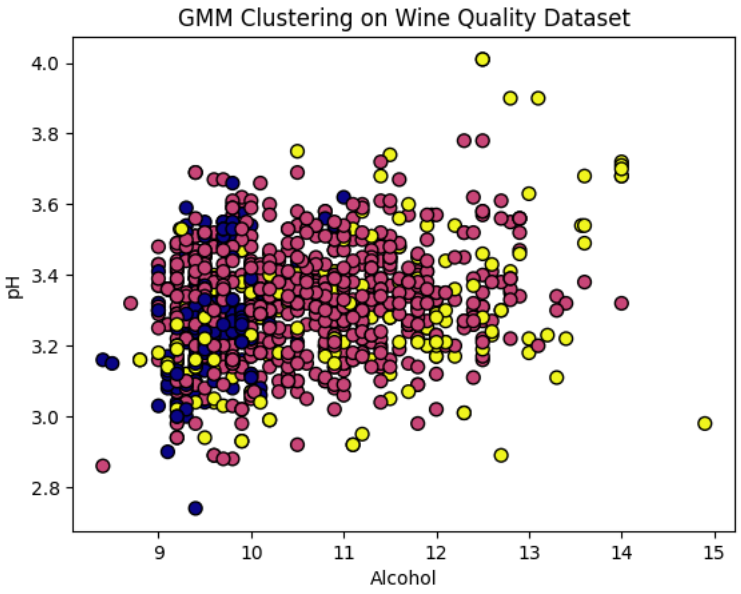
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality	Id	Cluster
0	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5	0	1
1	7.8	0.880	0.00	2.6	0.098	25.0	67.0	0.99680	3.20	0.68	9.8	5	1	0
2	7.8	0.760	0.04	2.3	0.092	15.0	54.0	0.99700	3.26	0.65	9.8	5	2	0
3	11.2	0.280	0.56	1.9	0.075	17.0	60.0	0.99800	3.16	0.58	9.8	6	3	1
4	7.4	0.700	0.00	1.9	0.076	11.0	34.0	0.99780	3.51	0.56	9.4	5	4	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1138	6.3	0.510	0.13	2.3	0.076	29.0	40.0	0.99574	3.42	0.75	11.0	6	1592	1
1139	6.8	0.620	0.08	1.9	0.068	28.0	38.0	0.99651	3.42	0.82	9.5	6	1593	1
1140	6.2	0.600	0.08	2.0	0.090	32.0	44.0	0.99490	3.45	0.58	10.5	5	1594	1
1141	5.9	0.550	0.10	2.2	0.062	39.0	51.0	0.99512	3.52	0.76	11.2	6	1595	1
1142	5.9	0.645	0.12	2.0	0.075	32.0	44.0	0.99547	3.57	0.71	10.2	5	1597	1

```
df.Cluster.value_counts()
```

Cluster	count
1	743
0	211
2	189

dtype: int64

```
# Step 5: Visualize clusters in 2D using two features ('alcohol' and 'pH')
plt.scatter(df['alcohol'], df['pH'], c=cluster_labels, cmap='plasma', edgecolor='k', s=50)  #viridis = colormap
#You can try other colormaps like 'plasma', 'inferno', 'magma', or 'cividis' depending on your preference.
plt.xlabel('Alcohol')
plt.ylabel('pH')
plt.title('GMM Clustering on Wine Quality Dataset')
plt.show()
```




```
#Step 6
# Print log-likelihood, BIC, and AIC for evaluation
print(f"Log Likelihood: {gmm.score(X) * X.shape[0]:.2f}")
print(f"BIC: {gmm.bic(X):.2f}")
print(f"AIC: {gmm.aic(X):.2f}")
```

```
Log Likelihood: -1678.92
BIC: 4998.50
AIC: 3823.85
```

```
# Step 7.0: Print model means
print("Model Means:")
print(gmm.means_)
```

```
Model Means:
[[ 8.04879527  0.63676512  0.25247431  2.21813678  0.12010921 16.70347383
  74.92983968  0.99697585  3.2726023   0.62832893  9.60889784]
 [ 8.36158259  0.50074922  0.26043209  2.16205704  0.07701776 14.50842789
  33.2576539   0.99658617  3.32322382  0.64687158 10.52666459]
 [ 8.40946661  0.53005127  0.315105    4.24762774  0.0875991  18.53918894
  61.2898545   0.99699842  3.30753668  0.72991749 11.0356997 ]]
```

Cluster	Fixed Acidity	Volatile Acidity	Citric Acid	Residual Sugar	Chlorides	Free SO_2	Total SO_2	Density	pH	Sulphates	Alc
1	8.05	0.64	0.25	2.22	0.12	16.70	74.93	0.9970	3.27	0.63	9.6
2	8.36	0.50	0.26	2.16	0.08	14.51	33.26	0.9966	3.32	0.65	10.5
3	8.41	0.53	0.32	4.25	0.09	18.54	61.29	0.9970	3.31	0.73	11.0

covariances

```
# Step 7.1: Print model covariances in a more readable format
print("\nModel Covariance for Cluster 1:")
cov_df1 = pd.DataFrame(gmm.covariances_[0], index=features, columns=features)
display(cov_df1)
```

Show hidden output

```
# Step 7.2: Print model covariances in a more readable format
print("\nModel Covariance for Cluster 2:")
cov_df2 = pd.DataFrame(gmm.covariances_[1], index=features, columns=features)
display(cov_df2)
```

Show hidden output

```
# Step 7.3: Print model covariances in a more readable format
print("\nModel Covariance for Cluster 3:")
cov_df3 = pd.DataFrame(gmm.covariances_[2], index=features, columns=features)
display(cov_df3)
```

```
Model Covariance for Cluster 3:
```

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
fixed acidity	5.173056	-0.077220	0.328928	1.287898	0.026814	-7.056401	-27.659680	0.004726	-0.299732	0.086473	-0.374764
volatile acidity	-0.077220	0.034500	-0.016932	-0.006099	0.000490	0.179656	0.088643	0.000078	0.011399	-0.008068	-0.061817
citric acid	0.328928	-0.016932	0.042274	0.102540	0.001685	-0.400692	-0.374028	0.000230	-0.023661	0.008712	0.024158
residual sugar	1.287898	-0.006099	0.102540	6.225811	0.022053	6.266380	7.472652	0.003753	-0.117169	-0.134651	-0.911037
chlorides	0.026814	0.000490	0.001685	0.022053	0.001241	0.095236	-0.158826	0.000047	-0.002230	0.001158	-0.016214
free sulfur dioxide	-7.056401	0.179656	-0.400692	6.266380	0.095236	174.598834	390.935575	-0.002077	0.060389	-0.281829	-4.561226
total sulfur	-27.659680	0.088643	-0.374028	7.472652	-0.158826	390.935575	1959.264277	-0.020114	-0.435038	0.896578	-7.655609