The objective of this assignment is to implement PCA on a given dataset and analyse the results.

In [1]:
```
import pandas as pd
df = pd.read_csv('wine.csv')
```

In [2]:
```
df
```

Out[2]:

	Wine	Alcohol	Malic.acid	Ash	Acl	Mg	Phenols	Flavanoids	Nonflavanoid.phenols
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39
...	...	...	...	...	...	...	...	...	...
173	3	13.71	5.65	2.45	20.5	95	1.68	0.61	0.52
174	3	13.40	3.91	2.48	23.0	102	1.80	0.75	0.43
175	3	13.27	4.28	2.26	20.0	120	1.59	0.69	0.43
176	3	13.17	2.59	2.37	20.0	120	1.65	0.68	0.53
177	3	14.13	4.10	2.74	24.5	96	2.05	0.76	0.56

178 rows × 14 columns

In [3]:
```
x = df.drop('Wine',axis = 1 )
y = df['Wine']
```

In [4]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.20 , ra
```

In [5]:
```
from sklearn.decomposition import PCA
pca=PCA(n_components=3)
pca
```

Out[5]:
```
PCA(n_components=3)
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

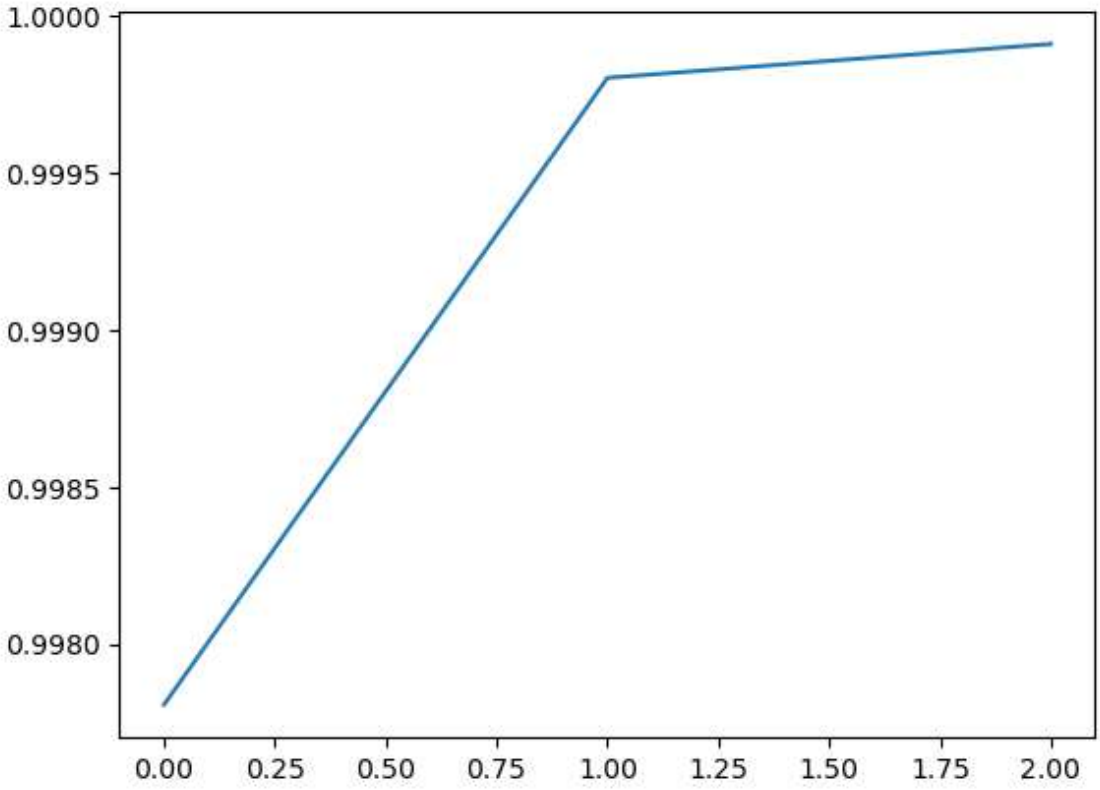
In [6]:
```
x_train=pca.fit_transform(x_train)
x_train
```

Out[6]:
```
array([[-7.49349515e+01, -8.62866599e-01, -7.23588298e+00],
       [-2.19971071e+02, -2.52597701e-01, -4.75918150e+00],
       [-7.51038966e+01, -1.10615564e+01,  1.58317882e+00],
       [-1.15087077e+02, -9.17867162e+00, -4.53814043e+00],
       [ 2.85024471e+02, -7.83445476e+00,  2.82671368e+00],
       [-4.22971985e+02, -6.02842277e-01,  3.30149652e+00],
       [-2.84862447e+02,  3.80522761e+00,  3.72465069e+00],
       [ 1.52899280e+01,  9.20255499e+00,  3.05842076e+00],
       [-1.85024331e+02, -5.01070474e+00, -3.63174215e-02],
       [ 4.50038279e+02, -7.81507097e+00, -1.44959905e+00],
       [-2.49481611e+01, -3.03617323e+00,  2.55125821e+00],
       [-3.69839061e+02,  8.54780405e+00, -4.03547114e+00],
       [-9.48291814e+01,  4.35146401e+00, -7.90255258e-01],
       [ 5.29900044e+02, -1.63311491e+01, -6.09145632e-01],
       [-3.43079063e+02, -8.13757163e+00,  5.01451248e+00],
       [ 2.35413393e+02,  1.31272425e+01,  2.01216202e+00],
       [ 4.59897892e+02, -1.60625383e+01,  9.51760726e-01],
       [-5.99146817e+01, -1.30757504e+00, -7.63503202e-01],
       [-1.55099663e+02, -9.56374993e+00, -7.23980592e-02],
```

In [7]:
```
pca.explained_variance_ratio_
```

Out[7]:
```
array([9.97808300e-01, 1.99598478e-03, 1.07916251e-04])
```

In [8]:
```
import numpy as np
from matplotlib import pyplot as plt
plt.plot(np.cumsum(pca.explained_variance_ratio_))
plt.show()
```



In [ ]:
```

```