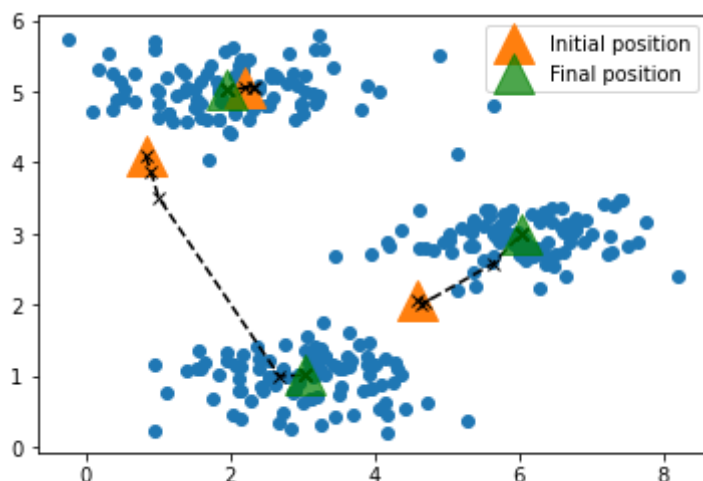
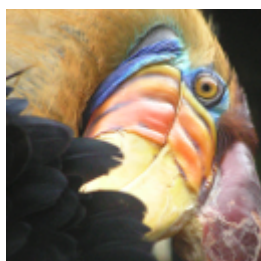


Below we show the result of running Kmeans algorithm on a 2d dataset. A random initial position for centroids is loaded. The figure below shows the movement of cluster centres(black dotted lines) with increasing number of iterations. Details of the dataset can be found in ex7.pdf.

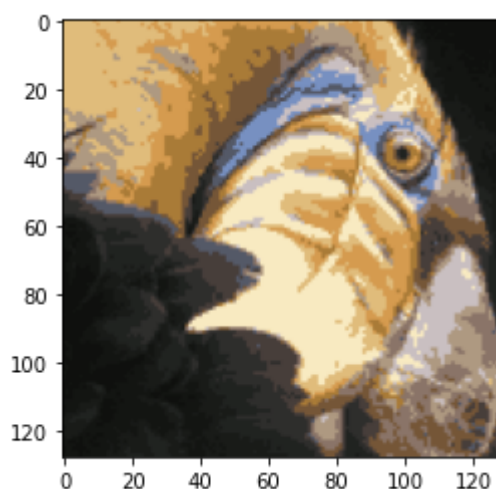
```
In [1]: import ex7_1
```



Next we consider an 128×128 pixels, RGB image. This is reshaped into a $128^2 \times 3$ dimensional array of training data for the colors. K-Means is applied to obtain 16 color centers and then each RGB value is replaced by the centroid color values it is closest to. The resulting image is shown below after the original image.



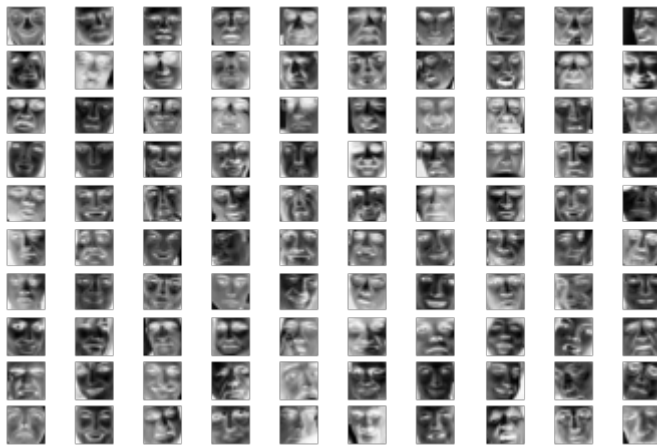
```
In [4]: %run ex7_2.py
```



Next we consider a dataset of face images, each 32×32 in greyscale. The first hundred faces are displayed below.

```
In [1]: import numpy as np
import scipy
import matplotlib.pyplot as plt
from PIL import Image
import scipy.io
```

```
In [2]: %matplotlib inline
file='ex7faces.mat'
tmp=scipy.io.loadmat(file)
X=tmp['X']
fig, axs = plt.subplots(10, 10)
for i in range(10):
    for j in range(10):
        tmp2=(X[10*i+j,:].reshape(32,32)).T
        axs[i,j].imshow(tmp2, cmap='gray_r')
        axs[i,j].axis('off')
```



We then perform a PCA on this dataset with 32^2 features and retain the first hundred principal components. The first hundred approximate face images reconstructed from this dimension reduced dataset is shown below.

```
In [3]: %run ex7_faces.py
```

