

# Programming Exercise 7: *K*-means Clustering and Principal Component Analysis

Machine Learning

## Introduction

In this exercise, you will implement the *K*-means clustering algorithm and apply it to compress an image. In the second part, you will use principal component analysis to find a low-dimensional representation of face images. Before starting on the programming exercise, we strongly recommend watching the video lectures and completing the review questions for the associated topics.

To get started with the exercise, you will need to download the starter code and unzip its contents to the directory where you wish to complete the exercise. If needed, use the `cd` command in Octave/MATLAB to change to this directory before starting this exercise.

You can also find instructions for installing Octave/MATLAB in the “Environment Setup Instructions” of the course website.

## Files included in this exercise

- `ex7.m` - Octave/MATLAB script for the first exercise on *K*-means
- `ex7_pca.m` - Octave/MATLAB script for the second exercise on PCA
- `ex7data1.mat` - Example Dataset for PCA
- `ex7data2.mat` - Example Dataset for *K*-means
- `ex7faces.mat` - Faces Dataset
- `bird_small.png` - Example Image
- `displayData.m` - Displays 2D data stored in a matrix
- `drawLine.m` - Draws a line over an existing figure
- `plotDataPoints.m` - Initialization for *K*-means centroids
- `plotProgresskMeans.m` - Plots each step of *K*-means as it proceeds

`runkMeans.m` - Runs the  $K$ -means algorithm  
`submit.m` - Submission script that sends your solutions to our servers  
[★] `pca.m` - Perform principal component analysis  
[★] `projectData.m` - Projects a data set into a lower dimensional space  
[★] `recoverData.m` - Recovers the original data from the projection  
[★] `findClosestCentroids.m` - Find closest centroids (used in  $K$ -means)  
[★] `computeCentroids.m` - Compute centroid means (used in  $K$ -means)  
[★] `kMeansInitCentroids.m` - Initialization for  $K$ -means centroids

★ indicates files you will need to complete

Throughout the first part of the exercise, you will be using the script `ex7.m`, for the second part you will use `ex7_pca.m`. These scripts set up the dataset for the problems and make calls to functions that you will write. You are only required to modify functions in other files, by following the instructions in this assignment.

## Where to get help

The exercises in this course use Octave<sup>1</sup> or MATLAB, a high-level programming language well-suited for numerical computations. If you do not have Octave or MATLAB installed, please refer to the installation instructions in the “Environment Setup Instructions” of the course website.

At the Octave/MATLAB command line, typing `help` followed by a function name displays documentation for a built-in function. For example, `help plot` will bring up help information for plotting. Further documentation for Octave functions can be found at the [Octave documentation pages](#). MATLAB documentation can be found at the [MATLAB documentation pages](#).

We also strongly encourage using the online **Discussions** to discuss exercises with other students. However, do not look at any source code written by others or share your source code with others.

---

## 1 $K$ -means Clustering

In this this exercise, you will implement the  $K$ -means algorithm and use it for image compression. You will first start on an example 2D dataset that

---

<sup>1</sup>Octave is a free alternative to MATLAB. For the programming exercises, you are free to use either Octave or MATLAB.

will help you gain an intuition of how the  $K$ -means algorithm works. After that, you will use the  $K$ -means algorithm for image compression by reducing the number of colors that occur in an image to only those that are most common in that image. You will be using `ex7.m` for this part of the exercise.

## 1.1 Implementing $K$ -means

The  $K$ -means algorithm is a method to automatically cluster similar data examples together. Concretely, you are given a training set  $\{x^{(1)}, \dots, x^{(m)}\}$  (where  $x^{(i)} \in \mathbb{R}^n$ ), and want to group the data into a few cohesive “clusters”. The intuition behind  $K$ -means is an iterative procedure that starts by guessing the initial centroids, and then refines this guess by repeatedly assigning examples to their closest centroids and then recomputing the centroids based on the assignments.

The  $K$ -means algorithm is as follows:

```
% Initialize centroids
centroids = kMeansInitCentroids(X, K);
for iter = 1:iterations
    % Cluster assignment step: Assign each data point to the
    % closest centroid. idx(i) corresponds to  $\hat{c}^{(i)}$ , the index
    % of the centroid assigned to example i
    idx = findClosestCentroids(X, centroids);

    % Move centroid step: Compute means based on centroid
    % assignments
    centroids = computeMeans(X, idx, K);
end
```

The inner-loop of the algorithm repeatedly carries out two steps: (i) Assigning each training example  $x^{(i)}$  to its closest centroid, and (ii) Recomputing the mean of each centroid using the points assigned to it. The  $K$ -means algorithm will always converge to some final set of means for the centroids. Note that the converged solution may not always be ideal and depends on the initial setting of the centroids. Therefore, in practice the  $K$ -means algorithm is usually run a few times with different random initializations. One way to choose between these different solutions from different random initializations is to choose the one with the lowest cost function value (distortion).

You will implement the two phases of the  $K$ -means algorithm separately in the next sections.

### 1.1.1 Finding closest centroids

In the “cluster assignment” phase of the  $K$ -means algorithm, the algorithm assigns every training example  $x^{(i)}$  to its closest centroid, given the current positions of centroids. Specifically, for every example  $i$  we set

$$c^{(i)} := j \quad \text{that minimizes} \quad \|x^{(i)} - \mu_j\|^2,$$

where  $c^{(i)}$  is the index of the centroid that is closest to  $x^{(i)}$ , and  $\mu_j$  is the position (value) of the  $j$ 'th centroid. Note that  $c^{(i)}$  corresponds to `idx(i)` in the starter code.

Your task is to complete the code in `findClosestCentroids.m`. This function takes the data matrix `X` and the locations of all centroids inside `centroids` and should output a one-dimensional array `idx` that holds the index (a value in  $\{1, \dots, K\}$ , where  $K$  is total number of centroids) of the closest centroid to every training example.

You can implement this using a loop over every training example and every centroid.

Once you have completed the code in `findClosestCentroids.m`, the script `ex7.m` will run your code and you should see the output `[1 3 2]` corresponding to the centroid assignments for the first 3 examples.

*You should now submit your solutions.*

### 1.1.2 Computing centroid means

Given assignments of every point to a centroid, the second phase of the algorithm recomputes, for each centroid, the mean of the points that were assigned to it. Specifically, for every centroid  $k$  we set

$$\mu_k := \frac{1}{|C_k|} \sum_{i \in C_k} x^{(i)}$$

where  $C_k$  is the set of examples that are assigned to centroid  $k$ . Concretely, if two examples say  $x^{(3)}$  and  $x^{(5)}$  are assigned to centroid  $k = 2$ , then you should update  $\mu_2 = \frac{1}{2}(x^{(3)} + x^{(5)})$ .

You should now complete the code in `computeCentroids.m`. You can implement this function using a loop over the centroids. You can also use a loop over the examples; but if you can use a vectorized implementation that does not use such a loop, your code may run faster.

Once you have completed the code in `computeCentroids.m`, the script `ex7.m` will run your code and output the centroids after the first step of  $K$ -means.

*You should now submit your solutions.*

## 1.2 $K$ -means on example dataset

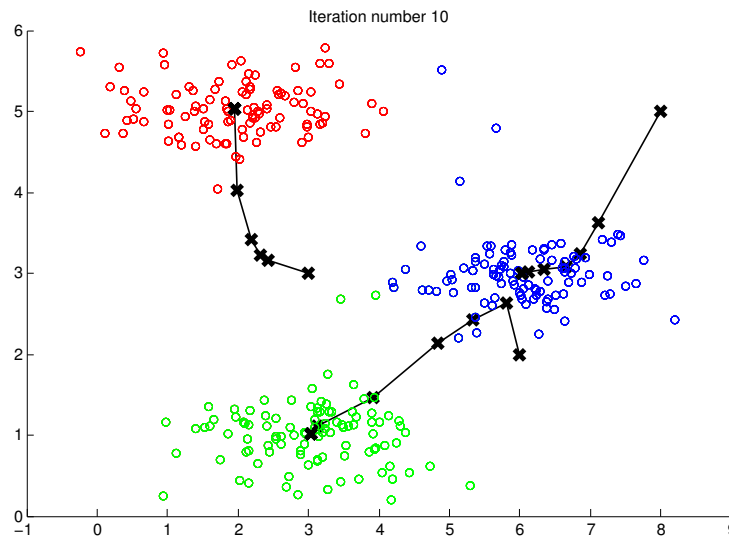


Figure 1: The expected output.

After you have completed the two functions (`findClosestCentroids` and `computeCentroids`), the next step in `ex7.m` will run the  $K$ -means algorithm on a toy 2D dataset to help you understand how  $K$ -means works. Your functions are called from inside the `runKmeans.m` script. We encourage you to take a look at the function to understand how it works. Notice that the code calls the two functions you implemented in a loop.

When you run the next step, the  $K$ -means code will produce a visualization that steps you through the progress of the algorithm at each iteration. Press *enter* multiple times to see how each step of the  $K$ -means algorithm changes the centroids and cluster assignments. At the end, your figure should look as the one displayed in Figure 1.

### 1.3 Random initialization

The initial assignments of centroids for the example dataset in `ex7.m` were designed so that you will see the same figure as in Figure 1. In practice, a good strategy for initializing the centroids is to select random examples from the training set.

In this part of the exercise, you should complete the function `kMeansInitCentroids.m` with the following code:

```
% Initialize the centroids to be random examples

% Randomly reorder the indices of examples
randidx = randperm(size(X, 1));
% Take the first K examples as centroids
centroids = X(randidx(1:K), :);
```

The code above first randomly permutes the indices of the examples (using `randperm`). Then, it selects the first  $K$  examples based on the random permutation of the indices. This allows the examples to be selected at random without the risk of selecting the same example twice.

*You do not need to make any submissions for this part of the exercise.*

### 1.4 Image compression with $K$ -means

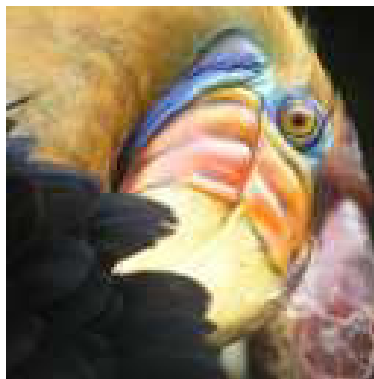


Figure 2: The original 128x128 image.

In this exercise, you will apply  $K$ -means to image compression. In a

straightforward 24-bit color representation of an image,<sup>2</sup> each pixel is represented as three 8-bit unsigned integers (ranging from 0 to 255) that specify the red, green and blue intensity values. This encoding is often referred to as the RGB encoding. Our image contains thousands of colors, and in this part of the exercise, you will reduce the number of colors to 16 colors.

By making this reduction, it is possible to represent (compress) the photo in an efficient way. Specifically, you only need to store the RGB values of the 16 selected colors, and for each pixel in the image you now need to only store the index of the color at that location (where only 4 bits are necessary to represent 16 possibilities).

In this exercise, you will use the  $K$ -means algorithm to select the 16 colors that will be used to represent the compressed image. Concretely, you will treat every pixel in the original image as a data example and use the  $K$ -means algorithm to find the 16 colors that best group (cluster) the pixels in the 3-dimensional RGB space. Once you have computed the cluster centroids on the image, you will then use the 16 colors to replace the pixels in the original image.

#### 1.4.1 $K$ -means on pixels

In Octave/MATLAB, images can be read in as follows:

```
% Load 128x128 color image (bird_small.png)
A = imread('bird_small.png');

% You will need to have installed the image package to use
% imread. If you do not have the image package installed, you
% should instead change the following line to
%
%   load('bird_small.mat'); % Loads the image into the variable A
```

This creates a three-dimensional matrix  $A$  whose first two indices identify a pixel position and whose last index represents red, green, or blue. For example,  $A(50, 33, 3)$  gives the blue intensity of the pixel at row 50 and column 33.

The code inside `ex7.m` first loads the image, and then reshapes it to create an  $m \times 3$  matrix of pixel colors (where  $m = 16384 = 128 \times 128$ ), and calls your  $K$ -means function on it.

After finding the top  $K = 16$  colors to represent the image, you can now

---

<sup>2</sup>The provided photo used in this exercise belongs to Frank Wouters and is used with his permission.

assign each pixel position to its closest centroid using the `findClosestCentroids` function. This allows you to represent the original image using the centroid assignments of each pixel. Notice that you have significantly reduced the number of bits that are required to describe the image. The original image required 24 bits for each one of the  $128 \times 128$  pixel locations, resulting in total size of  $128 \times 128 \times 24 = 393,216$  bits. The new representation requires some overhead storage in form of a dictionary of 16 colors, each of which require 24 bits, but the image itself then only requires 4 bits per pixel location. The final number of bits used is therefore  $16 \times 24 + 128 \times 128 \times 4 = 65,920$  bits, which corresponds to compressing the original image by about a factor of 6.

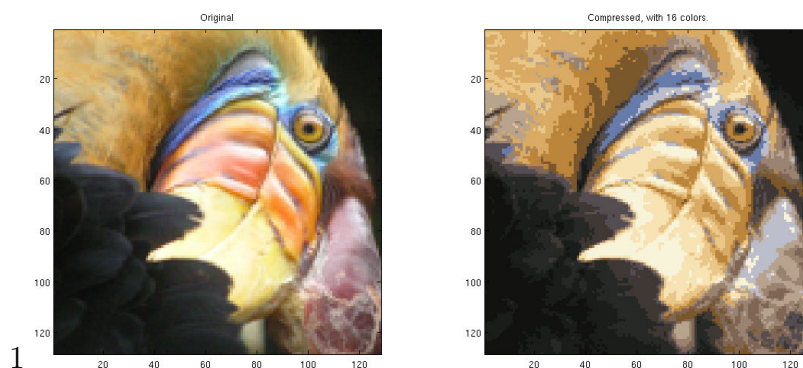


Figure 3: Original and reconstructed image (when using  $K$ -means to compress the image).

Finally, you can view the effects of the compression by reconstructing the image based only on the centroid assignments. Specifically, you can replace each pixel location with the mean of the centroid assigned to it. Figure 3 shows the reconstruction we obtained. Even though the resulting image retains most of the characteristics of the original, we also see some compression artifacts.

*You do not need to make any submissions for this part of the exercise.*

## 1.5 Optional (ungraded) exercise: Use your own image

In this exercise, modify the code we have supplied to run on one of your own images. Note that if your image is very large, then  $K$ -means can take a long time to run. Therefore, we recommend that you resize your images to manageable sizes before running the code. You can also try to vary  $K$  to see the effects on the compression.