



## Assignment 3 - K-Means Clustering

In this assignment you will implement K-Means clustering from scratch (That is, don't use a third-party machine learning implementation like scikit-learn; math libraries like numpy are fine).

### 1 Requirements

Go out and grab an image data set like:

- CIFAR-10 or CIFAR-100:  
<http://www.cs.toronto.edu/~kriz/cifar.html>
- MNIST Handwritten Digits:  
<http://yann.lecun.com/exdb/mnist/>
- Small NORB (toys):  
<https://cs.nyu.edu/~ylclab/data/norb-v1.0-small/>
- Street View Housing Numbers:  
<http://ufldl.stanford.edu/housenumbers/>
- STL-10:  
<https://cs.stanford.edu/~acoates/stl10/>
- Labeled Faces in the Wild:  
<http://vis-www.cs.umass.edu/lfw/>

Figure out how to load it into your environment and turn it into a set of vectors. Run K-Means on it for a few different K and show some results from the fit.

### 2 Deliverables

Submit a PDF report that includes the following:

- Well documented code snippets of your implementation.
- Run the algorithm for different values of K and provide the accuracy of the fit for each tried value with your comments.
- Are the results wildly different for different restarts and/or different K?
- What do the mean images look like? Provide screenshots with comments.



- What are some representative images from each of the clusters?
- Plot the K-Means objective function (distortion measure) as a function of iteration and verify that it never increases.

### 3 Notes

- You will work **individually**.
- Only send your report, don't send the code.

**Good Luck**