COEN 140L Lab 5 Intro

Xuyang Wu

Loading the data

from sklearn import datasets # if you do not have sklearn, please install it first. **pip install scikit-learn**
iris = datasets.load_iris()
print(list(iris.keys()))
print(iris.feature_names)
X = iris.data # each row is a sample
y=iris.target # target labels

K-means Clustering

```
# Initialize the cluster centers

for i in range(0,maxIter):
      # assignment, calculate the distance for each data point X
      # predicted labels, find the nearly center(class label)
      # update objective function J, for each cluster calculate the distance
      # stopping criterion check, check J from two iterations is below a certain value
      # update centers, calculate the mean and update the center
```

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{kn} \|\mathbf{m}_k - \mathbf{x}_n\|_2^2$$

Lab Tasks

- **Need demo** for week 5 assignment(10% points).
- Submit to Camino a pdf report with answers(60% points), the report contains some **results** which required by lab document, you also need to add some **observations** for the questions.
- Submit all the source code needed to generate these answers to Camino(30% points).