# Week 8

## Clustering

Unsupervised Learning: Introduction

* Supervised learning training set: {(*x*(1), *y*(1)), (*x*(2), *y*(2)), (*x*(3), *y*(3)), …, (*x*(*m*), *y*(*m*))}
* Unsupervised learning training set: {*x*(1), *x*(2), *x*(3), …, *x*(*m*)}
* Clustering algorithm: structure it finds is dataset grouped in clusters
* Applications of clustering

K-Means Algorithm

* Cluster assignment, move centroid
* Randomly initialize K cluster centroids (μ …)
* Repeat:
* Cluster assignment: c(i) := index of cluster centroid closest to x(i)? (value of k that minimizes the distance) min(k)||x(i) – μk||2
* Move centroid: μk := average(mean) of points assigned to cluster k
* Eliminate centroid with no points assigned, could randomly reinitialize
* Works for non-(well-)separated clusters 🡺 market segmentation

Optimization Objective