Exercises for Lecture 12

1. K-mean cost function

Remember the cost function for K-means method can be written as

$$J(r, \mu) = \sum_n \sum_k r_{nk} ||x_n - \mu_k||^2$$

where x_n denotes the nth training sample, μ_k denotes the center for cluster k, and r_n is one-hot cluster membership vector for x_n with $r_{nk} = 1$ for the cluster x_n belongs to.

We minimize the above cost function by repeating the following steps

1. Assign each sample to closest cluster

$$r_{nk} = 1 \quad \text{if } k = \text{argmin}_j ||x_n - \mu_j||^2$$

2. Set μ_k to the center of all samples in cluster k

$$\mu_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}}$$

Show that neither step 1 nor step 2 ever increase the cost J.

2. K-means lab

Do the labs in Section 10.5.1 in ISLR.