

# Homework

Jeff Tilton

August, 20 2019

## 1 Clustering

(a) Prove that using the squared Euclidean distance  $\|x_n - u_k\|^2$  as the dissimilarity function and minimizing the distortion function we have

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

That is,  $\mu_k$  is the center of the  $k$ -th cluster.

**Answer:**

The distortion function

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - u_k\|^2$$

is a quadratic function of  $\mu_k$ . It can be minimized by setting its derivative with respect to  $\mu_k$  to zero.

$$2 \sum_{n=1}^N r_{nk} (x_n - u_k) = 0$$

Solving for  $\mu_k$  gives the desired result

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

[1]

(B) Prove that K-means algorithm converges to a local optimum in finite steps.

**Answer:**

The K-means algorithm is an iterative algorithm with 2 steps after initialization of the first  $k$  clusters. The steps are repeated until convergence.

The steps are:

1. Minimize  $J$  with respect to  $r_{nk}$
2. Minimize  $J$  with respect to  $\mu_k$

The algorithm is monotonic, meaning the cost function decreases every iteration, which guarantees convergence.

(c) When we use the bottom-up hierarchical clustering to realize the partition of data, which of the three cluster distance metrics described above would most likely result in clusters most similar to those given by K-means? (Suppose  $K$  is a power of 2 in this case).

**Answer:**

Average linkage would separate the data most like the K-means algorithm.

Average linkage uses the average dissimilarity between groups [2]. This is essentially what K-means does as the first step of the algorithm. As the K-means algorithm minimizes  $J$  with respect to  $r_{nk}$  it is essentially comparing a cluster composed of a single point, to the average dissimilarity of all points in another cluster, represented by the cluster center. It then merges the cluster composed of a single point to whichever cluster center it is closest to.

(d) For the following data (two moons), which of these three distance metrics (if any) would successfully separate the two moons?

**Answer:**

Single linkage is the only distance metric that would successfully separate the two moons. Average and complete would misclassify some of the points where the moons overlap.

## References

- [1] Bishop, Christopher M. *Pattern recognition and machine learning*. Springer, 2006.
- [2] Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. Second Edition. *The elements of statistical learning*. Springer 2017.