

# CS 446: Machine Learning

## Homework

Due on Tuesday, April 3, 2018, 11:59 AM Central Time

### 1. [10 points] K-Means

- (a) Mention if K-Means is a supervised or an un-supervised method.

Your answer: K-Means is unsupervised method.

- (b) Assume that you are trying to cluster data points  $x_i$  for  $i \in \{1, 2, \dots, D\}$  into  $K$  clusters each with center  $\mu_k$  where  $k \in \{1, 2, \dots, K\}$ . The objective function for doing this clustering involves minimizing the euclidean distance between the points and the cluster centers. It is given by

$$\min_{\mu} \min_r \sum_{i \in D} \sum_{k=1}^K \frac{1}{2} r_{ik} \|x_i - \mu_k\|_2^2$$

How do you ensure hard assignment of one data point to one and only one cluster at a given time? Note: By hard assignment we mean that you are 100 % sure that a point either belongs or not belongs to a cluster.

Your answer:

We will choose  $r_{ik} \in \{0,1\}$  to be an indicator function:

$$r_{ik} = \begin{cases} 1, & \|x_i - \mu_k\|_2^2 \leq \|x_i - \mu_l\|_2^2 \quad \forall k \neq l \\ 0, & \text{otherwise} \end{cases}$$

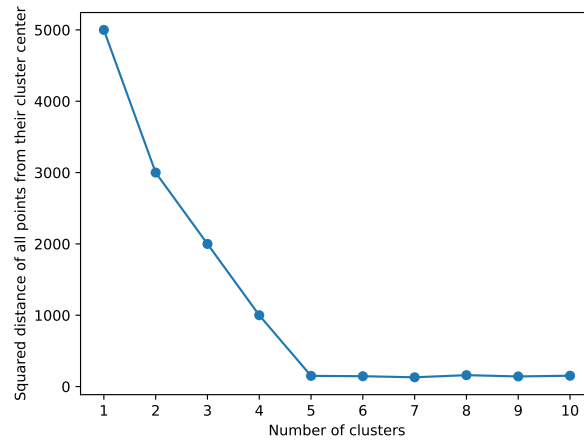
- (c) What changes must you do in your answer of part b, to make the hard assignment into a soft assignment? Note: By soft assignment we mean that you are sure that a point either belongs or not belongs to a cluster with some probability.

Your answer: Instead of using  $r_{ik} \in [0, 1]$ , This makes each row of matrix  $R$  a probability vector.

we can use something like soft max:

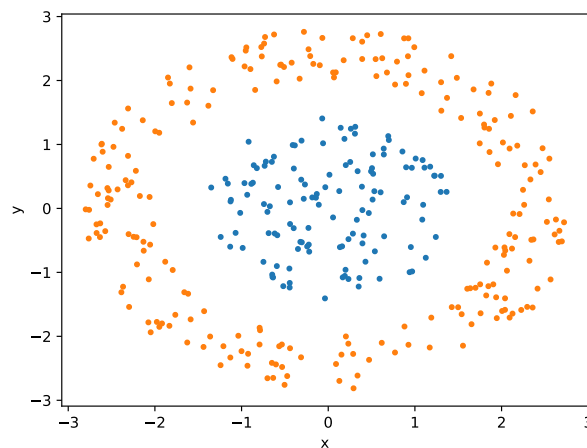
$$r_{ik} = \frac{\exp \|x_i - \mu_k\|_2^2}{\sum_{l=1}^K \exp \|x_i - \mu_l\|_2^2}$$

- (d) Looking at the following plot, what is the best choice for number of clusters?



Your answer: 5 clusters (elbow rule)

- (e) Would K-Means be an efficient algorithm to cluster the following data? Explain your answer in a couple of lines.



Your answer: No, K means does not take correlations into account. It finds the closest centroid for the assignment searching over spherically symmetric space. Hence, we will not observe outer (orange) shell as a separate cluster. What we will observe is a partition by a lines to clusters that have points from both blue ball and orange ring as members. Instead we should use spherical clustering for this.