

## 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:

- (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

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

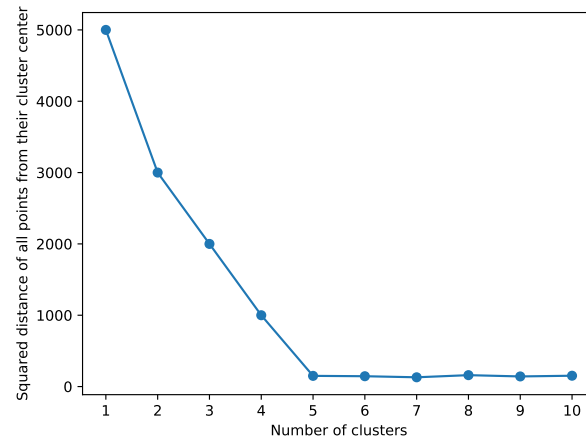
How do you ensure hard assignment of one data point to one and only one cluster at a given time? Note: By hard assignment we mean that you are 100 % sure that a point either belongs or not belongs to a cluster.

Your answer:

- (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.

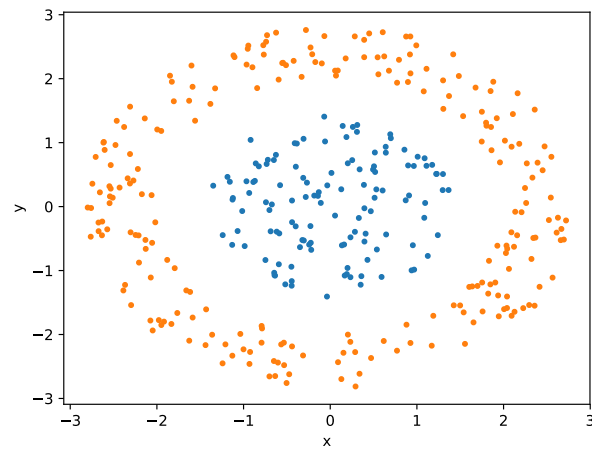
Your answer:

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



Your answer:

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



Your answer: