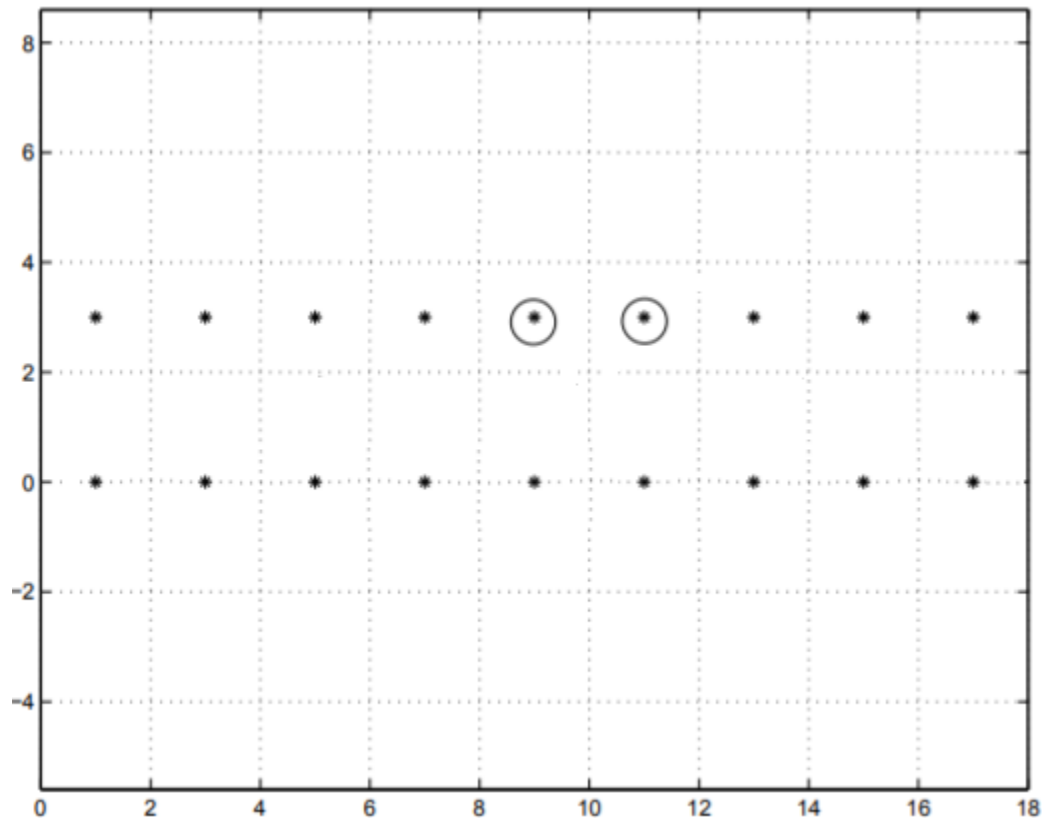# Clustering
Expectation-Maximization
Algorithm

Problem 1

Consider the data plotted below, which consist of two rows of equally spaced points. k-means clustering (k = 2) is initialized with the two points that are circled.

Indicate the final clusters obtained and their centroids.

Problem 2

Is it possible to get empty clusters during the k-means algorithm execution? If, so provide an example.

Problem 3

Can k-means algorithm be kernelized?

Problem 4

k-means algorithm is an iterative algorithm that finds the centers that minimize the sum of squared distances to the points.

Let's consider another objective function: maximizing spacing between clusters. We define the spacing of a set of clusters as the distance between the closest pair of points in *different* clusters. For a good clustering, we would expect the spacing to be large.

Design an algorithm to solve this problem.

Problem 5

Maximize the lower bound $F(\boldsymbol{\theta}, \{q_n\})$ over q_n assuming that $\boldsymbol{\theta}$ is fixed. See slide 46 in lecture 9.

Problem 6

On slide 48 in lecture 9 we defined a complete likelihood $Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{(t)}\right)$. Maximize Q over μ_k to get

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

Problem 7

There is a set of n points that we cluster using k-means and Gaussian mixtures. Suppose that in both cases we obtained 5 clusters and in both cases the centroids are the same.

Can 3 points that are assigned to different clusters in the k-means solution be assigned to the same cluster in the Gaussian mixture solution?