

Data Mining II
Quiz
March 10th

- 1) In k-means clustering **name and BRIEFLY describe two strategies** for selecting the number of clusters.

2) From ISLR:

In this problem, you will perform K -means clustering manually, with $K = 2$, on a small example with $n = 6$ observations and $p = 2$ features. The observations are as follows.

Obs.	X_1	X_2
1	1	4
2	1	3
3	0	4
4	5	1
5	6	2
6	4	0

- (a) Plot the observations.
- (b) Randomly assign a cluster label to each observation. You can use the `sample()` command in **R** to do this. Report the cluster labels for each observation.
- (c) Compute the centroid for each cluster.
- (d) Assign each observation to the centroid to which it is closest, in terms of Euclidean distance. Report the cluster labels for each observation.
- (e) Repeat (c) and (d) until the answers obtained stop changing.
- (f) In your plot from (a), color the observations according to the cluster labels obtained.