Nearest Neighbour Classifier

- Read 3 files for red, green, and unknown data sets
- For each unknown sample in the unknown data set
  - Calculate distances from the unknown sample to all red data samples
  - Find min_1 (minimum distance of the above distances to red samples)
  - Calculate distances from the unknown sample to all blue data samples
  - Find min_2 (minimum distance of the above distances to blue samples)
  - Compare min_1 and min_2 and assign class label to the unknown sample
- Output all unknown samples and their class label to screen
- Output all unknown samples and their class label to file

K-means Clustering

- Read data file, get number of dimensions D and number of data samples N
- Input number of clusters K, create K clusters same dimension D at random, and set threshold to a small value
- Repeat the following:
  - For each data sample, find its nearest cluster centre
  - Group data samples having the same nearest centre to a cluster
  - For each cluster, calculate new cluster centre (average of all samples)
  - Calculate sum of distances between old and new cluster centres
  - If the sum is less than the threshold: display K cluster centres and data samples on canvas then break, else: set cluster centres to new cluster centres