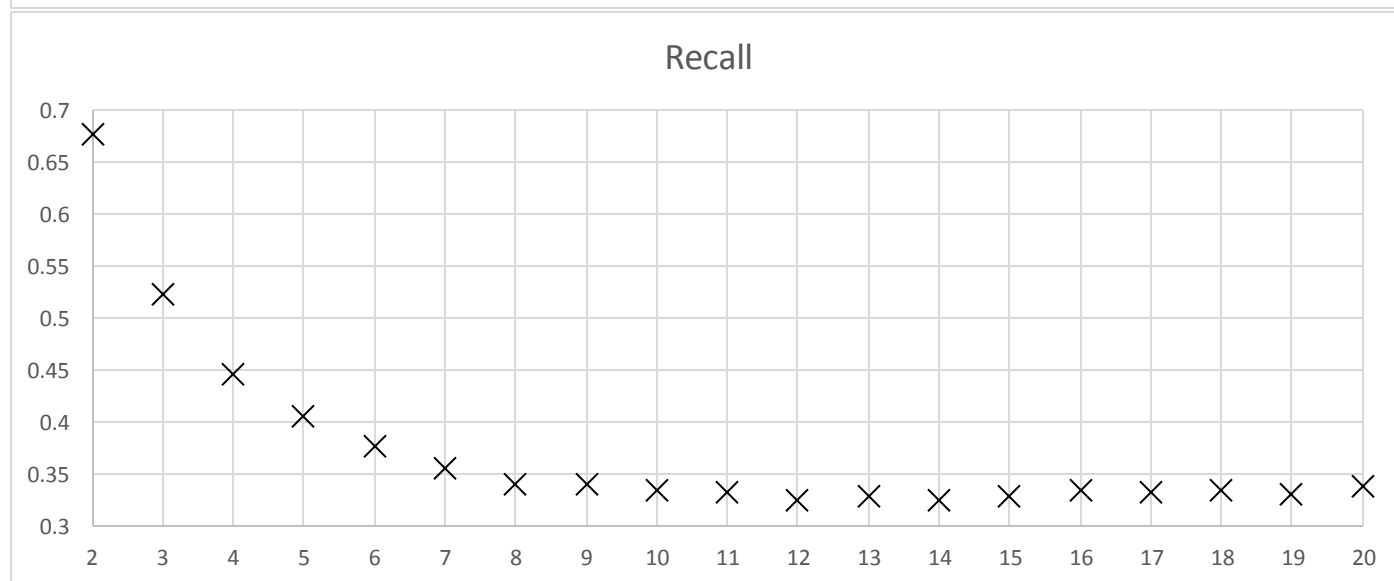
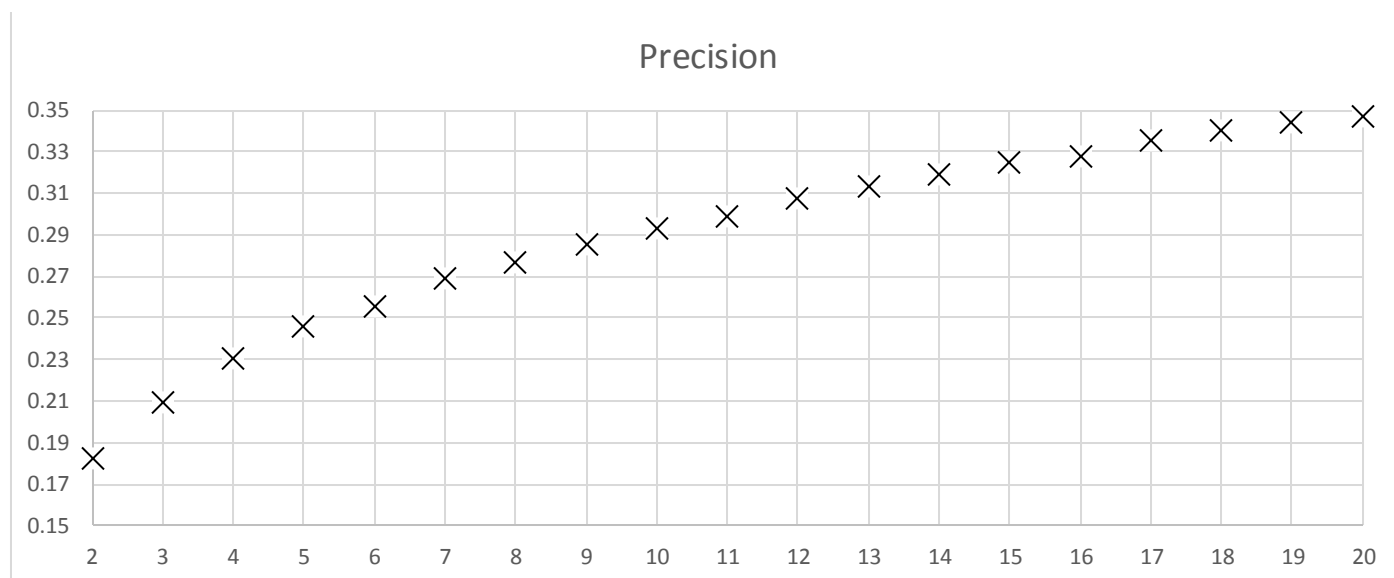
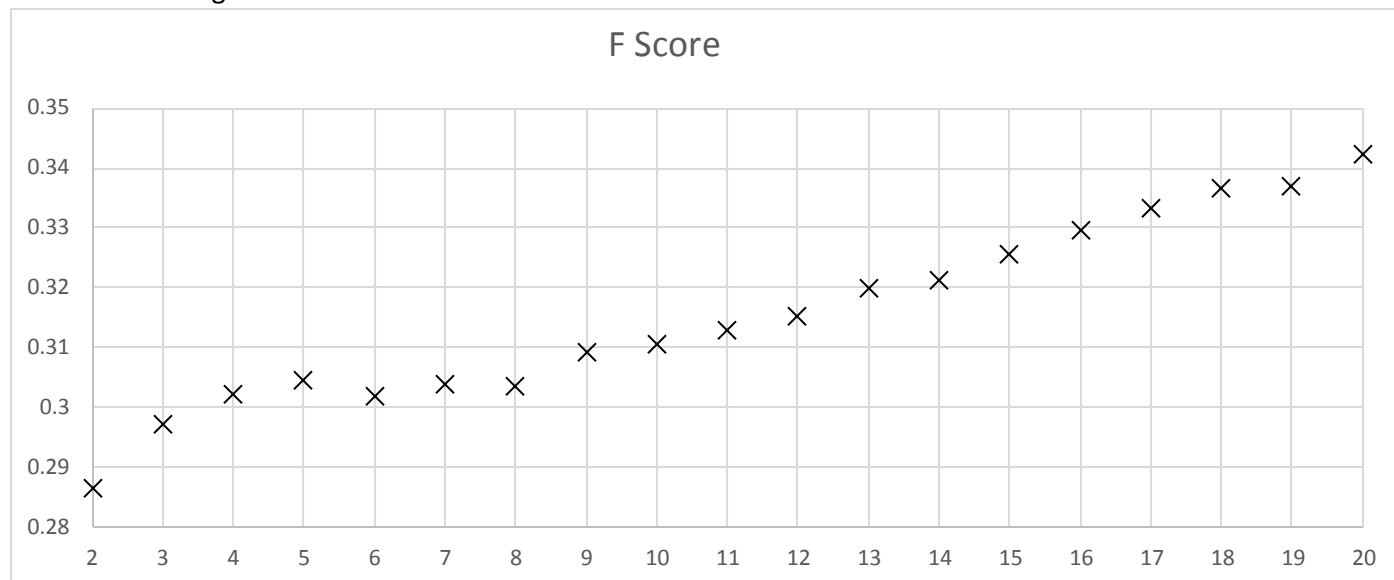
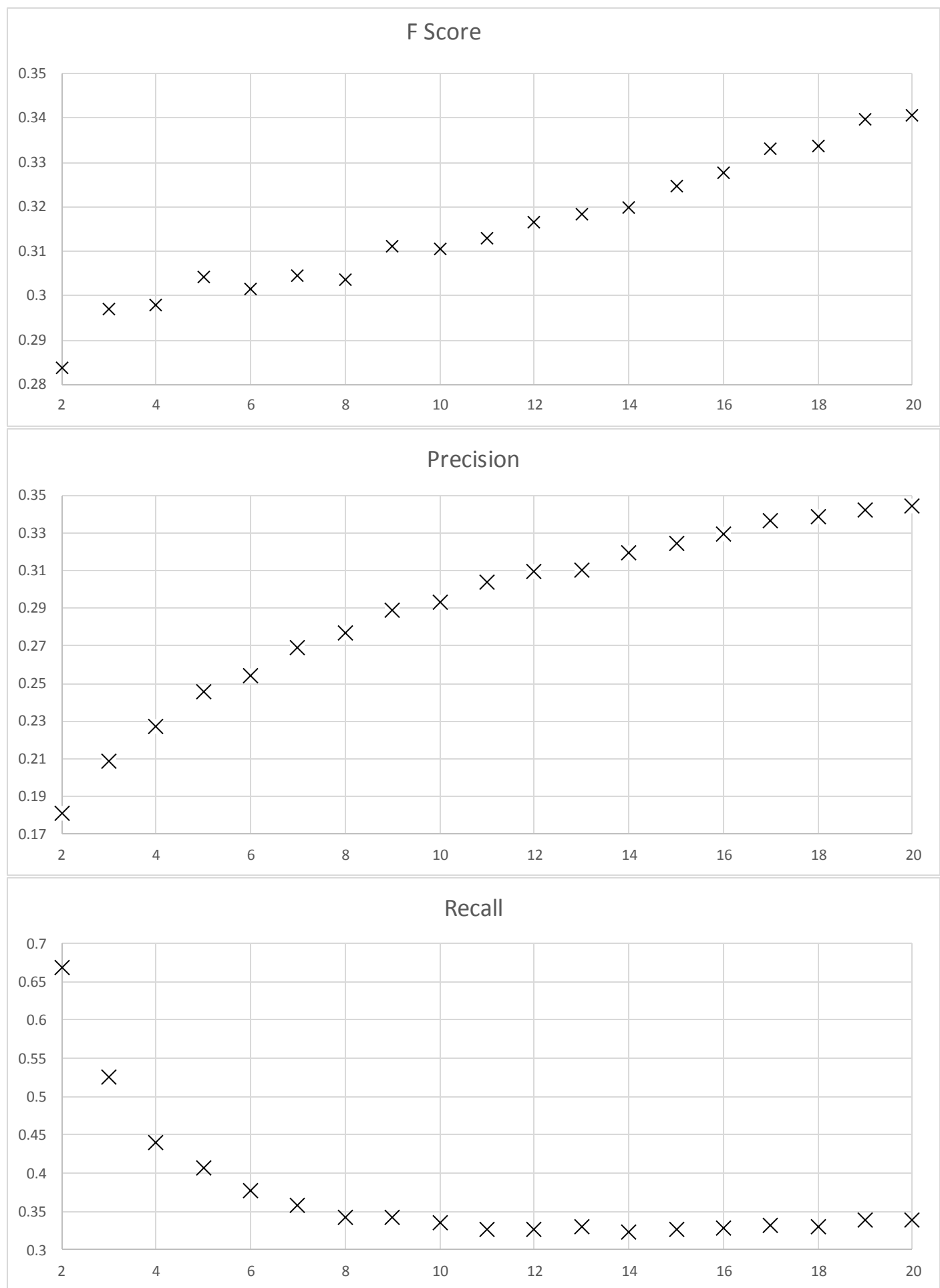


F Score, precision and recall calculated for k values 2-20, averaged over 200 iterations. The results can be reproduced using kmeans.py and data.txt.

Results for average cluster vector used as the centroid.



Instead of using the average vector of the cluster the datapoint nearest to the average can be used, this gives the following results.



With methods the recall drops off before stabilising at around 0.33, the precision steadily converges toward 0.33 while the f score rises very slowly while staying between 0.28 and 0.34.

The differences between the two centroid calculation methods are quite marginal but using the average vectors as a centroid produces a smoother graph due to it not being pegged to particular points where none suitable may exist.