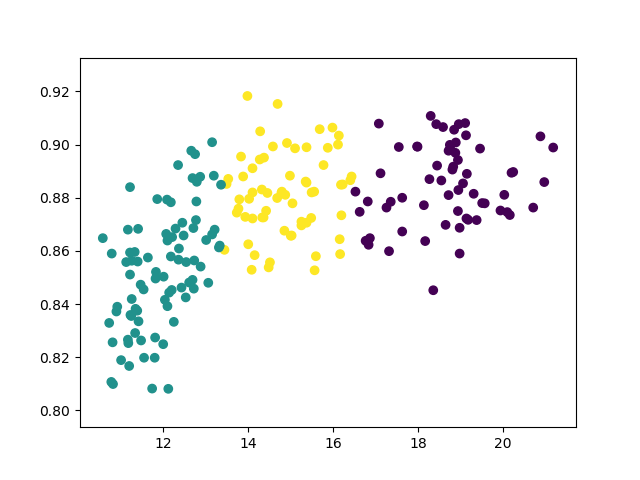
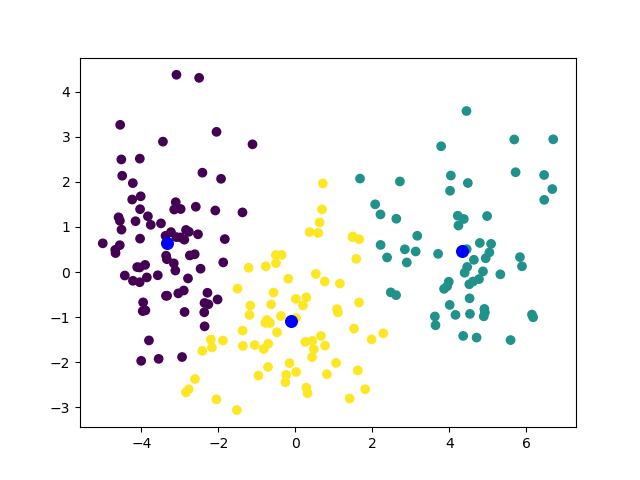
Report on Clusters – K-means and Gaussian mixture model.

To process our data, we had to use regex to spread the data with the delimiters tab and double tab. This becomes the data was not completely uniform, some columns had more tabs between them than others. But after we had done the data processing, we used NumPy and pandas to convert to a NumPy array.

In both models the clusters are quite spread out, with a lot of points from different clusters very close to each other. Our conclusion from this is that the difference between these kernels of seeds are relatively small and is therefore hard to separate. When we observed the raw data, we saw that many values had the same interval, independent of their class.

The visualization was decided on clearness and no points that were wrongly placed, i.e. Intermixed. While the boundaries between the clusters were thin, they are still present in our visualization without any overlap.

The k-means model dimensionality was reduced using PCA, while the Gaussian is tweaked using the parameters described in the scikit-learn package.

Gaussian mixture model K-means model