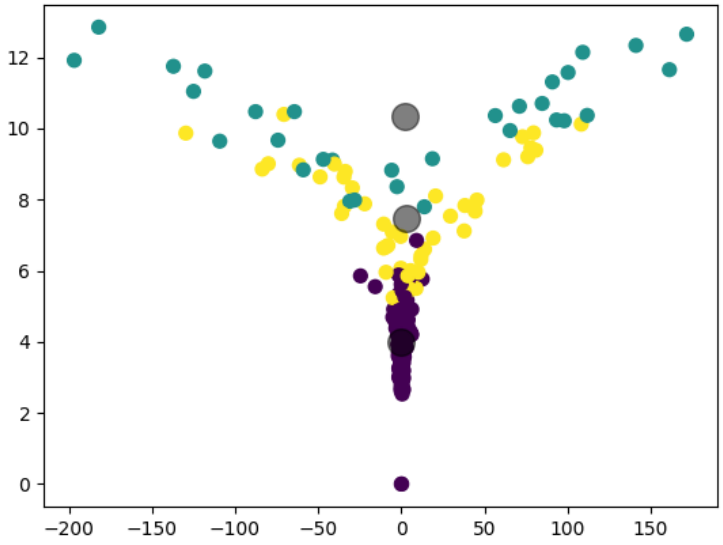
Phase2 – Evaluation with different classes

There are 5 different classes for this dataset. First, we create a three-class mode and test the performance of the previous steps algorithm on it. Then we create three two-class modes and test the algorithm of the previous steps on them.

- Three-class mode (AB-CD-E)

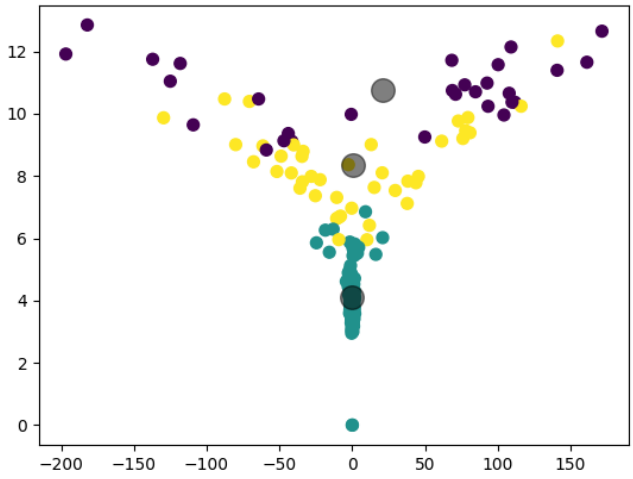
  Using k-means for clustering with different cluster numbers.

  - 3 clusters



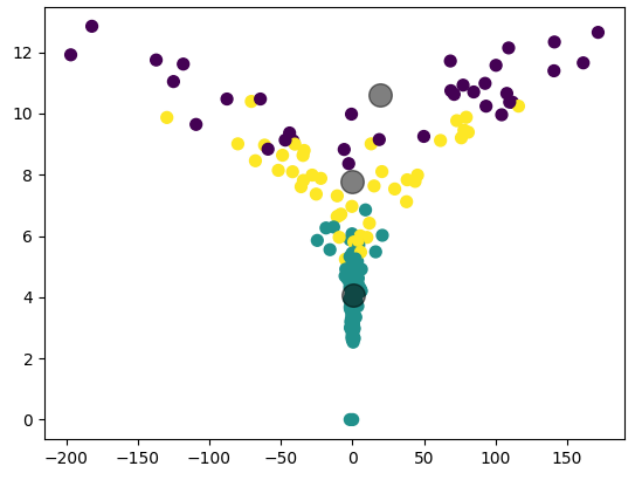
Out of 100 test data, it correctly recognizes 92 data. (0.92)

- Two-class mode (AB-E)
  - 3 clusters



Out of 100 test data, it correctly recognizes 60 data. (0.60)

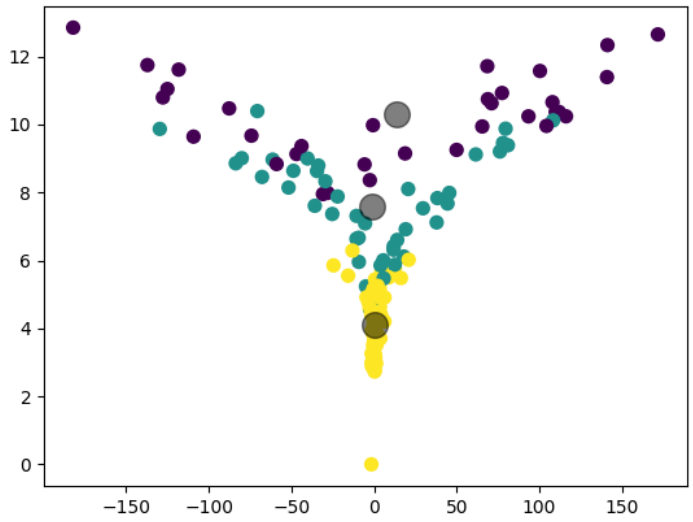
- Two-class mode (CD-E)
  - 3 clusters



Out of 100 test data, it correctly recognizes 60 data. (0.60)

- Two-class mode (D-E)

  - 3 clusters



Out of 100 test data, it correctly recognizes 39 data. (0.39)

- By choosing different classes from the dataset and new labeling, the accuracy decreases in general.
  In the case of 3 classes, because all datasets have been used and we have enough training data, the accuracy of recognizing 3 labels is still good.
  In the case of 2 classes especially the last one (D vs E) due to the reduction of training data and the use of more clusters than the number of classes, the accuracy is low.