Summary and Notes

# Exercise 4.1:

Best params according to our different experiments

window\_size = 25

threshold = 0.2

# Exercise 4.2:

It is sufficient to use posterior probability or a distance threshold as a single criterion. Both are used according to our dataset and goals. Both have their own advantages and disadvantages.

* Distance threshold is less sensitive to class imbalance, while Posterior Probability is more
* Posterior Probability handles ambiguity better than distance based.

# Exercise 4.3:

* The Clustering Algorithm is robust to initialization. Tested on 6 different seeds (see image).

# Exercise 4.4:

* Thresholds were found and plotted in the DIR Curve (see image).