Pattern Analysis Joint Meeting
Christian Riess

Worksheet 5: Tuesday May 18 / Friday May 21 2021

Please watch the video prior to the lecture, and think about the questions below. In the joint meeting, you will have 25 minutes time to discuss the questions with your group. Afterwards, we will jointly discuss your solution proposals.

You can print this sheet and use the space below for your notes.

Task 1: Mean Shift Clusters vs. k-Means/GMM Clusters

Following up on Task 2 from last week: on which inputs do you expect Mean Shift to produce noticeably different results from the k-Means and GMM algorithm? Assume that all three algorithms are parameterized such that the number of resulting clusters is approximately the same.

Task 2: Optimality Criterion in the Gap Statistics

Think about and explain in your own words what the optimality criterion (Eqn. (1) in Lecture 9) aims to achieve:

(a) What structural cue is hidden in the criterion $G(k) \leq G(k+1)$?

(b) Why the argmin?

(c) Let us also try a long shot, and (loosely) relate the optimality criterion to the training of a Density Forest

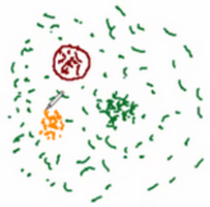
Task 3: Gap Statistics also for Mean Shift and GMMs?

It might be tempting to re-use the Gap Statistics for the Mean Shift and the GMM clustering to select the kernel size and the number of GMM components. However, I think that this is suboptimal. Which concerns could be raised here?

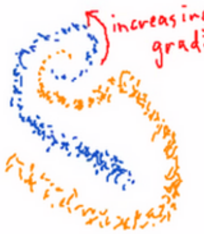
Task 1: Mean Shift Clusters vs. k-Means/GMM Clusters

- samples that are far away from others will be allocated in their own cluster. k-Means would assign them to the nearest bigger cluster

Cluster shapes: we saw that k-means always does Voronoi cells, GMMs can generate non-convex clusters and also clusters that are surrounded by another cluster.



Mean Shift: a sample moves "upwards" on the gradient within the kernel window
⇒ we can obtain almost arbitrary cluster shapes, as long as there is always an increasing path on the gradient



(this is a theoretical consideration of course, in practice you will probably not have such wild clusters)

Task 2: Optimality Criterion in the Gap Statistics

How to find the optimal value of k in k-means

argmin_k { G(k) >= G(k+1) - s_(k+1) }

Generally, less clusters are preferred

Consider the uniform distribution as the worst distribution (structure-wise) to obtain a k-mean s clustering

The gap is a relative measure of how well we summarize the structure in our distribution co mpared to that worst-case distribution

We expect the gap to increase as long as we are "on the right track", i.e., as long as we sum marize more structure better with adding clusters

If we add a cluster beyond the optimum number k, we "lose ground" with respect to the wor st case distribution, i.e., the gap decreases again

Hence, we choose k* as that k before the gap decreases minus some safety margin expressed by the standard deviation

On c): This is maybe remotely similar to the Density Forest, where we split the density with growing tree depth. Howev er, one difference is that we don't necessarily stop to grow the tree, but instead prune unnecessary splits back in a sec ond step