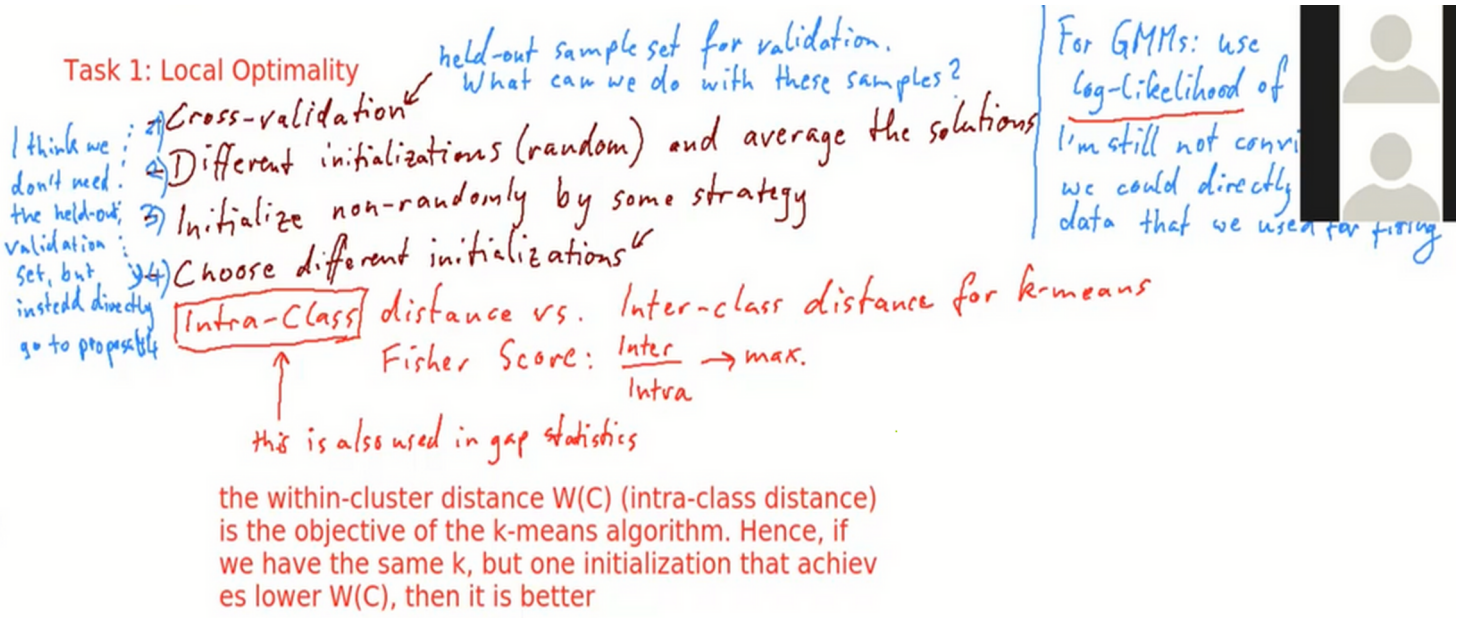


Task 1: Local Optimality

Both k-means and GMM fitting are only locally optimal, i.e., they find a solution that is only optimal w.r.t. the starting parameters. Assume that we have a customer who really wants to use k-means or GMMs, but local optima should be avoided.

How can we increase our chances to find a solution close to the global optimum? Can we propose a metric to determine whether solution A is closer to the global optimum than solution B?

Task 2: Clustering Forensics

K-means and GMMs may result in different clusterings. Think about the differences between the k-Means and GMM objective functions, and construct an input where they should give easily distinguishable results (for "reasonable" starting parameters).

Task 3: GMMs for Density Estimation

Since the GMM estimates a PDF $p(\mathbf{X})$, it could also be used for density estimation.

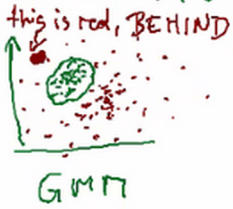
- How does the GMM density estimator relate to our other methods?
- For which sample distributions might a GMM density estimator be preferable?

Task 2: k-means: each point has a unique cluster (→ ML afer GMM also gives that on the other hand, you could also simulate probabilities in k-means)

k-means clusters can not overlap

Cluster boundary: GMM: "ellipsoid" more or less
k-means: Voronoi tesselation: lines bound the cluster

Isolated samples / substructures / other special things:



Task 3: Gaussian of course ☺
Gaps between clusters are filled
k-NN, kernel DE, Density Forests

GMM ⟷ Gaussian kernel
no objective function
↓
requires mean, cov.

separation of components
⇓
Estimates mean, covariance

Note some similarity in the resulting models of GMMs and Dens. Forests:
Both distribute a couple of Gaussians in the space

GMM: parametric, number of components needed

k-NN, Parzen wind.: Non-parametric "less prior knowledge required"

Density Forest: Parametric or non-parametric?

Each tree of height h gives ~ 2^h Gaussians

You can in principle let the trees grow to infinity, and prune them back to some "natural" size