Foundations of data science, summer 2020

JONATHAN LENNARTZ, MICHAEL NÜSKEN, ANNIKA TARNOWSKI

10. Exercise sheet
Hand in solutions until Thursday, 25 June 2020, 12:00

Exercise 10.1 (Mix of Gaussians). (10+2 points)

Consider a mixture of spherical Gaussians with density

$$f = w_0 p_0 + \cdots + w_{k-1} p_{k-1}$$

where $p_i = \mathcal{N}(|\mu_i\rangle, \sigma_i^2)$ for given vectors $|\mu_i\rangle \in \mathbb{R}^d$ and variance $\sigma_i^2 \in \mathbb{R}_{>0}$.

We want to experiment with projections.

For testing purposes use $k = 3$ and

$$|\mu_0\rangle = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, |\mu_1\rangle = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, |\mu_2\rangle = \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}$$

and $\sigma_0 = \frac{3}{4}$, $\sigma_1 = \sigma_2 = \frac{1}{4}$.

(i) Sample $N = 500$ points for each Gaussian. Put all $3N$ points as rows in a matrix A. [2]

(ii) Project all points to the x-y-plane and plot the result. What do you observe? [2]

(iii) Project all points to the y-z-plane and plot the result. What do you observe? [2]

(iv) Determine each centroid $|c_i\rangle$ and compare to $|\mu_i\rangle$. [2]

(v) Compute the singular value decomposition of A. [2]

Can you identify the (affine) subspace spanned by the $|c_i\rangle$?

(vi) Center the data, ie. determine the overall centroid and subtract it from each row. Determine the singular value decomposition of the resulting matrix. [+2]

Can you identify the (affine) subspace spanned by the $|c_i\rangle$?