Exercise 1

Elements of Machine Learning, 2022, by Jens Petersen

General comments

The assignments in EML must be completed individually and written individually. Group discussions are allowed and encouraged, but in such cases you should list the group members in your handin. Your handin should include your solution to the exercises in two-files, a report in pdf format (not zipped), describing your solution, and a zip-file including all the relevant code. Each part to the exercise counts equally towards the overall score/grade.

Exercises

Solve the following exercises in PRML[1] (10 points each)

- 8.3, 8.4, 8.10, 9.5

Old Faithful

For this exercise you will need to download the Old Faithful data set from `https://www.stat.cmu.edu/~larry/all-of-statistics/=data/faithful.dat`

1. Load and plot the data with similar x and y axes as Figure 2.21 in PRML (10 points).
2. Implement and describe your implementation of the EM algorithm. Use the implementation to estimate two mixtures (mean and covariance) in the data. Hint: you may need to initialize the covariance matrix with random numbers or some other strategy (10 points)
3. Plot the points and marginal probabilities for each component. You could for instance color the points or the background by marginal probability for each component (10 points).
4. What happens if you use more than two mixtures? Try to reason about your observations. (10 points).

[1] Pattern Recognition and Machine Learning, Christopher M. Bishop