685.621 Algorithms for Data Science
Homework 4
Assigned at the start of Module 7
Due at the end of Module 8
Total Points 100/100

Collaboration groups have been set up in Blackboard. Make sure your group starts one thread for the collaborative problem. You are required to participate in the collaborative problem. Do not directly post a complete solution, the goal is for the group to develop a solution after everyone has participated.

1. **Problem 1**
   20 Points Total

   In this problem, develop pseudocode and code for the Expectation Maximization method. This should be done for a generic number of clusters, at a minimum you should be able to handle 3 clusters to build a three class classifiers. Using the following data

$$\mathbf{x} = \begin{bmatrix} 1 & 2 \\ 4 & 2 \\ 1 & 3 \\ 4 & 3 \end{bmatrix} \tag{1}$$

   for 5 iteration show the values for $p^(i)(k|n), \mu_k^(i+1), \sigma_k^(i+1), p_k^(i+1)$ using your code. You can either use a built in EM algorithm or the one you implement to show how well the clusters create the two separations as in slide 15 of the Expectation Maximization.pdf for the 5 iterations. In this example, are the clusters starting to converge? If no, why not? If yes, why?

2. **Problem 2** ***Note this is a Collaborative Problem***
   20 Points Total

   Using the EM algorithm from Problem 1 on the IRIS data set estimate the the unknown parameters μ_k, σ_k, p_k.

3. **Problem 3**
   20 Points Total

   Consider three mean values of $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3] = [4.5, 2.2, 3.3]$ with a corresponding covariance matrix as follows:

$$\Sigma = \begin{bmatrix} 0.5 & 0.1 & 0.05 \\ 0.1 & 0.25 & 0.1 \\ 0.05 & 0.1 & 0.4 \end{bmatrix} \tag{2}$$

   The respective minimums are min $= [3.5, 1.7, 2.5]$ and maximums are max $= [5.5, 2.7, 4.1]$.

   Generate 300 observations.

   Using the EM algorithm from Problem 1 and the generated date estimate the the unknown parameters μ_k, σ_k, p_k.

4. **Problem 4**
   40 Points Total

   Using the IRIS data set, the EM algorithm and the Bayes classifier, generate a three class classification method. Show your results for the three class case. Provide analysis on the results generated.

References

[1] Bishop, Christopher M., *Neural Networks for pattern Recognition*, Oxford University Press, 1995

[2] Bishop, Christopher M., *Pattern Recognition and Machine Learning*, Springer, 2006, https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

[3] Duin, Robert P.W., Tax, David and Pekalska, Elzbieta, *PRTools*, http://prtools.tudelft.nl/

[4] Dempster, A. P., Laird, N. M. and Rubin, D. B., *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society B, Volume 39, Number 1, pp.1–22, 1977

[5] Franc, Vojtech and Hlavac, Vaclav, *Statistical Pattern Recognition Toolbox*, https://cmp.felk.cvut.cz/cmp/software/stprtool/index.html

[6] Fukunaga, Keinosuke, *Introduction to Statistical Pattern Recognition*, Academic Press, 1972

[7] Machine Learning at Waikato University, *WEKA*, https://www.cs.waikato.ac.nz/ ml/index.html

[8] Tomasi, C., *Estimating Gaussian Mixture Densities with EM – A Tutorial*, Duke University Course Notes, 2006, http://www.cs.duke.edu/courses/spring04/cps196.1/handouts/EM/tomasiEM.pdf, Retrieved Sept 2006