Homework 1

CS420 Machine learning 2018 Spring*
Department of Computer Science and Engineering
Shanghai Jiao Tong University

Submission deadline: 20:00, April 12, 2018, Thursday

Submission to:

Please submit your homework in pdf format to the CS420 folder in the following FTP. File name should be like this: `015033910032_chenyajing_hw1.pdf`.

ftp://public.sjtu.edu.cn
username: cyj907
password: public

1 k-mean vs GMM

Give a variant of k-mean algorithm somewhat between the original k-mean and Expectation-Maximization (EM) for Gaussian Mixture Models (GMM). Please specify the computational details of the formulas. Pseudo-codes of the algorithm would be great.

Discuss the advantages or limitations of your algorithm.

2 k-mean vs CL

Compare the k-mean algorithm with competitive learning (CL) algorithm. Could you apply the idea of Rival Penalized Competitive Learning (RPCL) to k-mean so that the number of clusters is automatically determined? If so, give the details of your algorithm and then implement it on a three-cluster dataset generated by yourself. If not, state the reasons.

3 model selection of GMM

Write a report on experimental comparisons on model selection performance between BIC,AIC and VBEM.

Specifically, you need to randomly generate datasets based on GMM, by varying some factors, e.g., sample sizes, dimensionality, number of clusters, and so on.

- BIC, AIC: First, run EM algorithm on each dataset X for $k = 1, ..., K$, and calculate the log-likelihood value $\ln[p(X|\hat{\Theta}_k)]$, where $\hat{\Theta}_k$ is the maximum likelihood estimate for parameters; Second, select the optimal k^* by

$$k^* = \arg\max_{k=1,...,K} J(k), \tag{1}$$

$$J_{AIC}(k) = \ln[p(X|\hat{\Theta}_k)] - d_m \tag{2}$$

*tushikui@sjtu.edu.cn

$$J_{BIC}(k) = \ln[p(X|\hat{\Theta}_k)] - \frac{\ln N}{2} d_m \tag{3}$$

- Use VBEM algorithm for GMM to select the optimal k^* automatically or via evaluating the lower bound.

The following codes might be useful.

Matlab: `http://www.cs.ubc.ca/~murphyk/Software/VBEMGMM/index.html`

Python: `http://scikit-learn.org/stable/modules/mixture.html`