

Comments on Assignment 3

Jian Yao

CSC 321 : Introduction to Neural Networks and Machine Learning
Department of Computer Science
University of Toronto

March 10, 2014

Part 1

- The training and validation dataset are drawn from 5 random **axis aligned** Gaussians. (What if not axis aligned)
- In the training, you are supposed to change the number of gaussians in your model and plots both the negative log validation density as well as the training density as a function of the number of gaussians. Use a standard deviation of 0.1 in the plot. Please specify these numbers in the plot or put these numbers in a separate table. (3 points)
- Give a brief statement of what you think of the graph. (1 point)
- Give a brief statement about the effects of changing the initial standard deviation used in 'mogem.m'. (1 point)

Stop Criterion

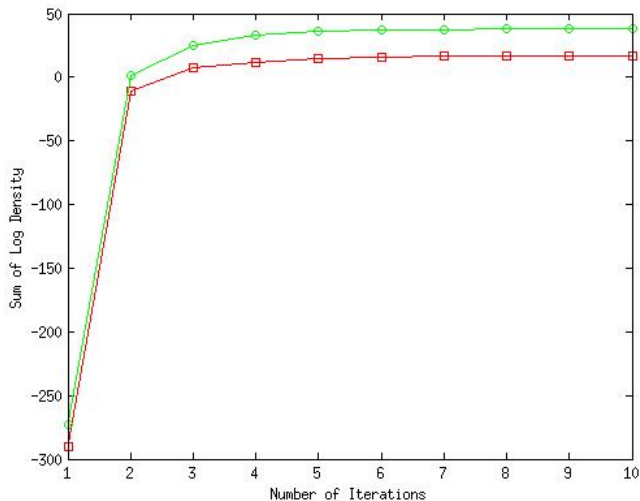
In the function 'mogem', the training stops when the number of iterations reaches 'numupdates' you set at the beginning. Is it good? Why?

Stop Criterion

In the function 'mogem', the training stops when the number of iterations reaches 'numupdates' you set at the beginning. Is it good? Why?

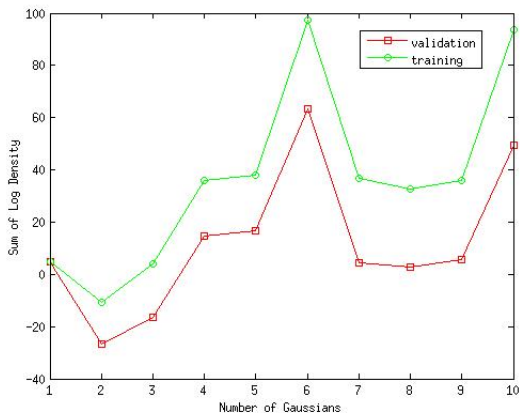
If you set 'numupdates' to be very large, then the training converges but waste a lot of time which is not necessary! Here's one suggestion: once you observe the objective function in training changes little in the update, then stop it. You can also stop when you observe the mean/variance changes little.

Stop Criterion



example of the plot in Part 1

This example might not be correct since I did not consider whether the training converges!!!!



- Present the graph; (1 point)
- Your observation and reasoning; (1 point)

What is the optimal number of gaussians in training? Does the observation coincide with what you think? Why?

Part 3

If the mixing proportions change, the posterior distribution of the latent variable will also be changed. So in this part, you have to rewrite the expression of the posterior distribution in 'mogem.m'.

If the mixing proportions change, the posterior distribution of the latent variable will also be changed. So in this part, you have to rewrite the expression of the posterior distribution in 'mogem.m'.

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)}$$

If the mixing proportions change, the posterior distribution of the latent variable will also be changed. So in this part, you have to rewrite the expression of the posterior distribution in 'mogem.m'.

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)}$$

Print out the final mixing proportions(start with the 4 settings provided in the assignment) and report in your submission (3 points)

Thanks!