

## HW7 (ECE7252 Spring 2008)

1. Exercise 6.2.
2. Suppose we have a linear function  $y=f(x)$  applied to data uniformly spread over the interval  $[0, 10]$ , but we fit a kernel regression function such as the Nadaraya-Watson kernel-weighted average in Equation (6.2). Which part of the interval in  $[0, 10]$  will be most prone to error? Why?
3. A random vector  $\mathbf{x}$  has a mixture Gaussian density as shown in Eq. (6.32), derive its expected mean vector and covariance matrix.
4. The data set hw7-gmm.txt is generated by a Gaussian mixture model. Find a Matlab routine that perform the EM algorithm described in Lecture Note 17-18 to estimate the unknown parameters, mixture gains, means and variances. You can first plot the histogram of the data. Then you can try to fit 2, 3, 4 and 5 mixture components, and compare the overall likelihoods of the training data in each case. Remember an EM is an iterative procedure that will often converge to a local minimum solution that maximize the overall likelihood locally, plot in each case the overall likelihood as a function of the number of iteration. Set a stopping criterion for your iteration process and state why you choose such a criterion. You are required in some cases to choose some initial guess of the parameter estimates in order to start the iteration. State your reason for your choice.

**PS. This homework was done in a rush. If you found some problem too hard, maybe I have overlooked something. You will be given credit if you show your effort even you have not solved the given problem.**

**NB. Quiz 1 will cover all the materials we have covered so far in the course with more emphasis on the later parts on ANN, SVM and kernel methods. Some review will be done during the week right after the spring break, and right before the quiz.**