

Exercise 4, Oct. 12, 2006

1. The associative Gaussian mixture model (the probability density function generated by the Mixture of Experts) was presented in the scalar case. Generalize the model to the multivariate case in which the desired response is a vector of dimension q and the error is a multivariate Gaussian distribution with zero mean and covariance matrix Σ . How does the generalization change the network?
2. Derive the stochastic gradient algorithm for the training of the mixture of experts model.
3. (a) Construct the block diagram of an HME model with three levels of hierarchy. Assume the use of a binary decision tree for the model.
(b) Formulate the conditional probability density function for the HME model described in part (a)
4. Prove that the EM algorithm works. *Hint:* see Problem 7.10 in Haykin.