COURSE 02457

Non-Linear Signal Processing: Exercise 9

This exercise is based on in C. M. Bishop: *Neural Networks for Pattern Recognition* sections 2.6, 5.6-5.7.

Print and comment on the figures produced by the software as outlined below at the **Checkpoints**.

Regression based on density estimation

We observe a stochastic multi-channel signal with inputs $\mathbf{x}$ and outputs y and our aim is model the density $p(\mathbf{x}, y) \sim p(\mathbf{x}, y|\mathbf{w})$ where the family $p(\mathbf{x}, y|\mathbf{w})$ is a given parametric density. The training set is consists of N input-output pairs.

The gaussian mixture model family is defined

$$p(\mathbf{x}, y|\mathbf{w}) = \sum_{j=1}^{M} p(\mathbf{x}, y|j, \mathbf{w}_j)P(j).$$

Where each component density is a normal distribution. Here we will invoke a family of "isotropic" Gaussians, i.e., Gaussians with covariance matrices that are scaled unit matrices,

$$p(\mathbf{x}, y|\boldsymbol{\mu}_j, \sigma_j^2) = \frac{1}{(2\pi\sigma_j^2)^{(d+1)/2}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu}_j)^2 + (y-\nu_j)^2}{2\sigma_j^2}\right).$$

We will estimate the parameters using the Expectation-Maximization algorithm as in exercise 8 (see also Bishop page 67).

Based on the joint input-output distribution we can compute the conditional output mean $y(\mathbf{x}) =< t|\mathbf{x} >$ (Bishop pages 178-179),

$$y(\mathbf{x}) = \frac{\sum_{j=1}^{M} \frac{P(j)\nu_j}{(2\pi\sigma_j^2)^{\frac{d}{2}}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu}_j)^2}{2\sigma_j^2}\right)}{\sum_{j'=1}^{M} \frac{P(j')}{(2\pi\sigma_{j'}^2)^{\frac{d}{2}}} \exp\left(-\frac{(\mathbf{x}-\boldsymbol{\mu}_{j'})^2}{2\sigma_{j'}^2}\right)}$$

Checkpoint 9.1

Use the program `main9a.m` to perform EM learning of the parameters for the sunspot data. The program creates four figures. The first figure is a scatter plot of training points $x(k)$ versus $y(k) = x(k+1)$, and also the location of the centers as iterations progress. The second plot shows the evolution of the variances σ_j. By default we clamp them to be identical by setting `common-sigs = 1`. The second figure also shows the evolution of the training and test errors of the density estimates. The third figure is like figure 1, but with the final clusters. Figure 4 shows the training and test set time series. The program also reports (in the Matlab prompt line) the training and test errors estimated by prediction. These are calculated as in ealier exercises by the normalized squared error of the predictions on training and test sets. In the program you can also change the number of mixture components K. First run the program with $K = 5, 25$ with common covariances. Do you see overfitting in the density test error (figure 2)? Set `common-sigs = 0` and comment on the results for $K = 5, 25$. If you experience problems with shrinking variances, try to introduce a lower bound, what is a good value? Compare the results obtained with those of multilayer perceptron nets from earlier exercises.

Checkpoint 9.2

To run the program `main9a.m` with higher k values you may want to eliminate intermediate plots by setting `plot-motion=0`. Run the algorithm for higher K's to find the best K from a prediction point of view.

Signal detection based on density estimation

If we adapt density models $P(\mathbf{x}|C_k)$ for each class of a classification problem we can use Bayes' theorem to obtain the posterior probabilities (see Bishop pages 179-181),

$$P(C_k|\mathbf{x}) = \frac{\sum_j p(\mathbf{x}|j)P(j|C_k)P(C_k)}{\sum_{j'} p(\mathbf{x}|j')P(j')}$$

Checkpoint 9.3

Use the program `main9b.m` to adapt densities for the two classes of the "pima indian problem". In figure 1 we show scatter plot of two of the seven pima inputs and superimpose the motion of the centers as we adapt the densities. This program also reports two test errors: a density estimate test error (figure 2) and the classification test error (prompt line). Figure 3 shows the location of the components for the final configuration. Create a flowchart description of the program. Run the program for different K, comment on the location of the components. Which value of K can you recommend?

DTU, 2006 Lars Kai Hansen