Exercises in Doctoral Course in Speech Recognition

Fall semester 2003

The solutions shall be reported along with the results. They can be in the form of handwritten or typed notes, Excel diagrams, or computer programs (preferably C or Java). The author's solutions have been written in Excel.

1. VQ quantisation. Build a VQ codebook consisting of four codewords using the K-means algorithm and the LBG algorithm. Use Euclidean distortion metric ($d=(x-\mu_x)^2+(y-\mu_y)^2$). End criterion: Distortion decrease between iterations is less than 0.1 or number of iterations is at least 3. Report the centroids of the codewords.
   The 16 samples of 2-dimensional input data for building the codebook are:
   {3.4, 4.9}, {8.1, 4.3}, {6.8, 7.2}, {5.9, 2.0}, {9.9, 3.7}, {2.8, 6.0}, {4.0, 1.3}, {7.7, 8.6},
   {8.2, 6.8}, {3.3, 6.7}, {7.7, 2.3}, {9.6, 2.0}, {4.7, 6.6}, {6.9, 4.0}, {1.9, 5.5}, {4.9, 2.4}
   The initial codewords are: {0.0, 5.0}, {2.0, 3.0}, {3.0, 2.0}, {5.0, 0.0}

2. CART algorithm Decision tree.
   The same 16 samples as in Ex. 1. have been assigned with three feature categories f1, f2 and f3 in the following way:
   {X, Y, f1, f2, f3} =
   { {3.4, 4.9, 0, 0, 1},
   {8.1, 4.3, 1, 1, 0},
   {6.8, 7.2, 1, 1, 1},
   {5.9, 2.0, 1, 0, 0},
   {9.9, 3.7, 1, 1, 1},
   {2.8, 6.0, 0, 0, 1},
   {4.0, 1.3, 0, 0, 0},
   {7.7, 8.6, 1, 1, 1},
   {8.2, 6.8, 1, 1, 1},
   {3.3, 6.7, 0, 1, 1},
   {7.7, 2.3, 1, 0, 0},
   {9.6, 2.0, 1, 1, 0},
   {4.7, 6.6, 1, 1, 1},
   {6.9, 4.0, 1, 1, 1},
   {1.9, 5.5, 0, 0, 1},
   {4.9, 2.4, 0, 0, 0} }

   Compute the first branch of a regression tree using these samples. The regression error in a node is defined as the weighted squared error of all samples from their node average:

   $$\overline{V}(t) = \left[\frac{1}{N(t)} \sum_{i=1}^{I} \left[(x_i - \mu_i)^2 + (y_i - \mu_i)^2 \right] \right] P(t)$$

   The question set is limited to independent checking of the value of each of the features. That is, three questions: f1 > 0? , f2 > 0? , f3 > 0? .
   Use the splitting criterion

   $$\Delta\overline{V}(q) = \overline{V}(t) - (\overline{V}(l) + \overline{V}(r))$$

   Which question is used in the branch? What are the regression errors before and after that split? How many samples are in the two new nodes?

3. A certain process is modelled by an HMM consisting of three states, s_1, s_2 and s_3. Their initial state occupance probabilities are {0.5, 0.3, 0.2}, respectively. The transition probability matrix is:

   $$\begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$$

The observation symbol inventory is {a, b, c, d} and the observation probabilities for the states are:

$$\begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \end{array}\begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} 0.4 & 0.1 & 0.2 & 0.3 \\ 0.3 & 0.1 & 0.5 & 0.1 \\ 0.2 & 0.3 & 0.1 & 0.4 \end{bmatrix} \end{array}$$

At one occasion, the observed output symbol sequence of the process was:
{b, a, c, c, d}
Compute the Viterbi and Forward probabilities between the observation sequence and the model.

4. Perform one iteration of the Baum-Welch re-estimation algorithm on the model in the previous example. The observations are used as training data. Report the new transition and observation probabilities and the probability that the re-estimated model has generated the observation sequence.

5. (Part of an example in the Pattern Recognition course at TMH,KTH)
A speech signal consisting of digits is described by one acoustic variable which is quantized to eight discrete values, ranging from 1 to 8. Our Markov model has as initial probability vector

$$\boldsymbol{\pi} = \begin{pmatrix} 0.3 & 0.1 & 0.1 & 0.3 & 0.1 & 0.1 \end{pmatrix}$$

as transition matrix

$$A = \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.0 & 0.5 & 0.4 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 \\ 0.2 & 0.0 & 0.0 & 0.4 & 0.0 & 0.4 \end{bmatrix}$$

and the densities of the observations for each state are described by

$$\mathbf{B} = \begin{bmatrix} 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.2 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.2 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.4 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.6 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.3 & 0.2 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.7 & 0.3 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 \end{bmatrix}$$

The observation sequence is

$$\mathbf{o} = \begin{pmatrix} 5 & 5 & 5 & 6 & 6 & 7 & 7 & 5 & 5 & 6 & 6 & 7 & 1 & 1 & 3 & 3 & 4 & 4 \end{pmatrix}$$

(a) How many digits do you think the model describes and why?

(b) By looking at the model specification and the sequence, try to predict the optimal digit sequence.