Exercises in Doctoral Course in Speech Recognition

Spring semester 2007

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:
   {1.1, 6.9}, {1.5, 8.3}, {2.0, 4.2}, {2.5, 1.9}, {3.3, 5.2}, {3.9, 3.6}, {4.2, 5.1}, {4.6, 6.3}, {5.1, 9.3}, {5.4, 7.6}, {5.6, 3.4}, {7.4, 1.5}, {7.5, 4.2}, {7.6, 8.1}, {8.3, 0.6}, {8.6, 2.6}
   The initial codewords are: {3.0, 3.0}, {7.0, 3.0}, {3.0, 7.0}, {7.0, 7.0}

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} =
   { {1.1, 6.9, 0, 0, 1},
   {1.5, 8.3, 1, 1, 0},
   {2.0, 4.2, 1, 1, 1},
   {2.5, 1.9, 1, 0, 0},
   {3.3, 5.2, 1, 1, 1},
   {3.9, 3.6, 0, 0, 1},
   {4.2, 5.1, 0, 0, 0},
   {4.6, 6.3, 1, 1, 1},
   {5.1, 9.3, 1, 1, 1},
   {5.4, 7.6, 0, 1, 1},
   {5.6, 3.4, 1, 0, 0},
   {7.4, 1.5, 1, 1, 0},
   {7.5, 4.2, 1, 1, 1},
   {7.6, 8.1, 1, 1, 1},
   {8.3, 0.6, 0, 0, 1},
   {8.6, 2.6, 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. The probability density function of the frame-level observation vector in HMM- and GMM-based speech and speaker recognition systems is modeled as a mixture of Gaussian distributions according to the following formula:

$$b_j(y) = \sum_{m=1}^{M} c_{jm} b_{jm}(y); \qquad \sum_{m=1}^{M} c_{jm} = 1$$

where the emission probability of each mixture component b_{jm} is

$b_{jm}(y) = N(y; \mu_{jm}, \Sigma_{jm})$.

a. What is the likelihood value for one frame of an utterance when matched against a GMM model consisting of three mixture components? For computational simplicity, the input vector is 2-dimensional. The covariance matrix is diagonal.
Frame vector **y** = {7,12}
Mixture components *(format {mean and var of Dim 1}, { mean and var of Dim 2})* :
b_{j1} = {{4,3}, {9,21}}; b_{j2} = {{0,2}, {6,5}}; b_{j3} = {{11,30}, {18,73}}
Component weights: c_1 = 0.4; c_2 = 0.5, c_3 = 0.1
b. Estimate the order of magnitude of the total likelihood of one utterance of 100 frames under the (non-realistic) assumptions that they all have the same value as above and are statistically independent. What is a computer's value range of a 32-bit floating point number? What is the problem and what is normally done to avoid it?
c. How would the likelihood values change if the same distribution was modeled by, e.g. , 30 components instead of 3? Increase or decrease by a factor of 10, or remain unchanged?
d. In what direction will the likelihood value change by increasing the vector size?

4. 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 {0.4, 0.3, 0.3}, respectively. The transition probability matrix is:

$$\begin{array}{c} s_1 \\ s_2 \\ s_3 \end{array} \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.3 & 0.6 & 0.1 \\ 0.3 & 0.2 & 0.5 \end{bmatrix}$$

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

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

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.

5. 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.

M. Blomberg 2007-05-24