DT2118 Lab2: Hidden Markov Models with Gaussian Emissions

Mateusz Buda buda@kth.se
Masoumeh Poormehdi Ghaemmaghami mpg@kth.se

April 27, 2016

1 The models

The HMM models topology implies that they always start in the first state. This is encoded in the start probability vector that has 1 in the first position and 0s elsewhere. Moreover, the transition matrix only allows to stay in the same state or move one state forward.

2 Multivariate Gaussian Density

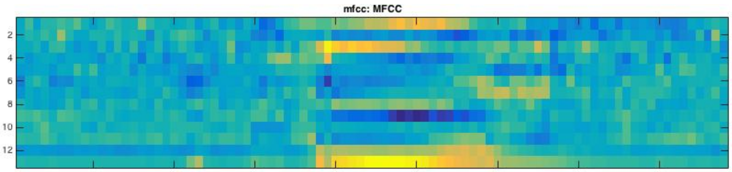

Figure 1: MFCC

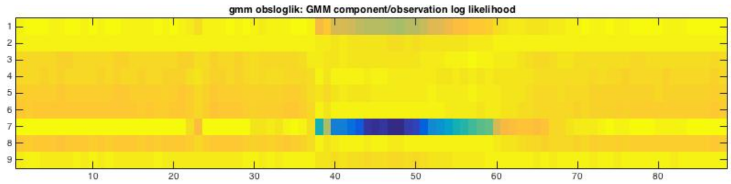

Figure 2: GMM component/observation log likelihood

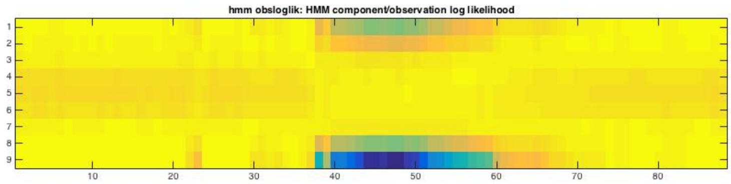

Figure 3: HMM component/observation log likelihood

In the figure 2 and 3 we can see that appropriate Gaussians get 'activated', i.e. have high values, in the time range of MFCC that is not silence.

3 GMM Likelihood and Recognition

Log likelihood of an observation sequence $\boldsymbol{X} = \{x_1, \ldots, x_N\}$ given the GMM model together with weights of each Gaussian density and assuming that the x_n vectors are independent for each $n = (0, N]$ is given by

$$\ln p(\boldsymbol{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{j=1}^{M} \pi_j \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right\}, \tag{1}$$

where π is a vector of weights [1]. Scoring each of the 44 utterances in the `tidigits` array with each of 11 GMM models and selecting the one with the highest log likelihood we recognized all 44 digits correctly.

4 HMM Likelihood and Recognition

4.1 Forward Algorithm

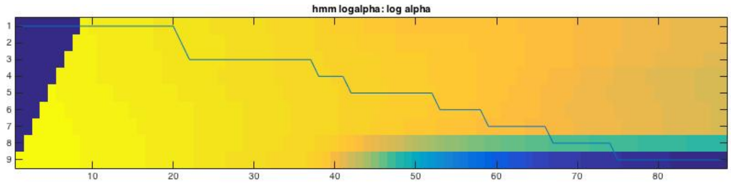

Figure 4: Log alpha

Figure 4 presents the output of `forward` function. Elements of this array correspond to

$$\log \alpha_n(j) = \log P(x_1, \ldots, x_n, z_n = s_j|\theta). \tag{2}$$

Based on that, the likelihood $P(\boldsymbol{X}|\theta)$ of the whole sequence $\boldsymbol{X} = \{x_1, \ldots, x_N\}$, given the model parameters θ and assuming that

$$P(x_1, \ldots, x_n, z_n = s_i, z_n = s_j|\theta) = 0, \quad \forall i, j, \quad i \neq j, \quad \textit{and}$$

$$\sum_{j=1}^{M} \alpha_n(j) = 1,$$

is given by

$$P(\boldsymbol{X}|\theta) = \sum_{i=1}^{M} P(\boldsymbol{X}, s_i|\theta) = \sum_{i=1}^{M} \alpha_N(i).$$

After conversion into log domain we have

$$\log P(\boldsymbol{X}|\theta) = \log \sum_{i=1}^{M} \exp(\log \alpha_N(i)). \tag{3}$$

Scoring each of the 44 utterances in the `tidigits` array with each of 11 HMM models using equation 3 and selecting the one with the highest log likelihood we have made 2 mistakes. This result is slightly worse than the one achieved by GMM but the error below 5% can be still considered satisfactory.

Using the Gaussian distributions in the HMM models as if they were GMM model with equal weights resulted in all 44 digits being recognized correctly. This is because for each frame we take into account all the Gaussians. With HMM transition model we select only those Gaussions related to a particular state. Current state is probabilistic and corresponds to the weights in the GMM model.

From the figure 4 we can read that in general HMM topology is pushing the probability forward. At the beginning most of the probability mass is concentrated mostly in the first states and in the end it is moved to the last ones. After silence ends, first two states loose their probability very fast.

4.2 Viterbi Approximation

Best path obtained by Viterbi decoding is drawn in the figure 4. It shows that we most likely stay in the same state that we were before or go to the next one. This reflects transition probabilities matrix of the HMM models.

Scoring each of the 44 utterances in the `tidigits` array with each of 11 HMM models and selecting the one with the highest Viterbi log likelihood we have made 2 mistakes. This is the same result that that one we obtained using equation 3 to compute log likelihood.

References

[1] Christopher M. Bishop. *Pattern Recognition and Machine Learning*, chapter Mixture Models and EM. Springer Science, New York, 2006.