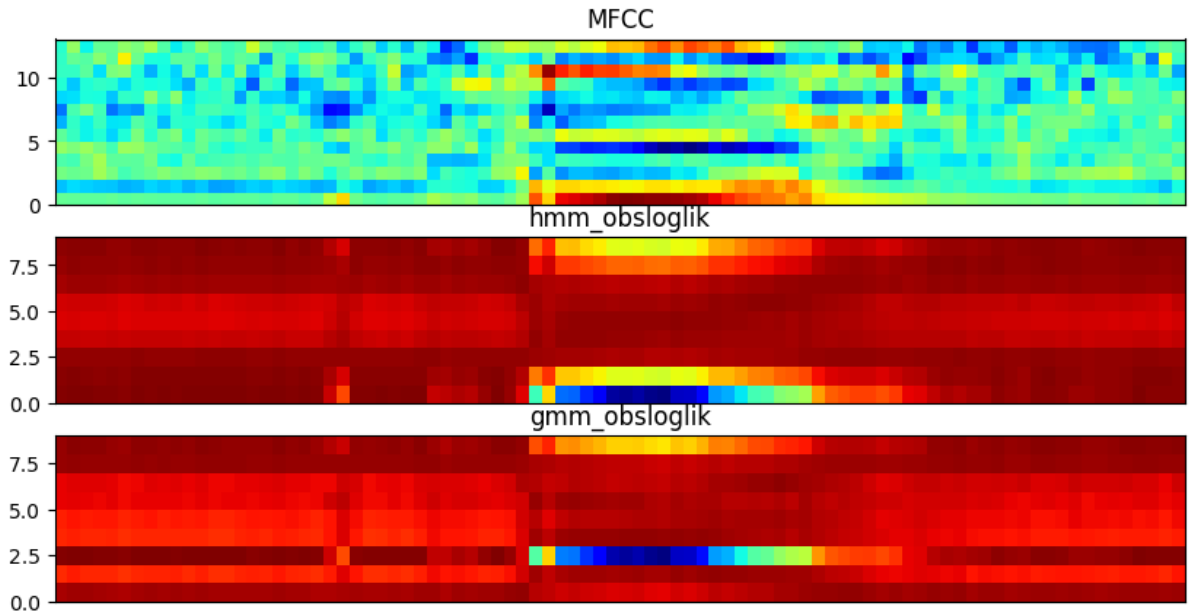
Speech and Speaker Recognition DT2119 "Lab2: Hidden Markov Models with Gaussian Emissions"

April 20, 2017

4 Multivariate Gaussian Density

In the case of a Gaussian Mixture Model (GMM) we compute the log likelihood of each observation xi (each frame of an utterance signal) and each Gaussian distribution. In the case of a Hidden Markov Model (HMM) we compute the emission probability of each observation and each state of the HMM.

Figure 1: log likelihood for Gaussians from HMMs and GMMs models



5 GMM Likelihood and Recognition

Here we compute the log likelihood of a sequence of ob servations (one utterance) for a specific GMM. That is, how probable it is that the sequence of observations is produced by a specific model.By calculating the likelihood of every utterance for every model, we essentially calculate the probability, that an utterance is indeed the digit of the corresponding model. Afterwards, we choose the model with the highest likelihood and we assign its digit to the utterance.

Figure 2: GMM Classification

```
------ GMM------
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 9 - 9
tid digit, mod digit: ---> 9 - 9
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 9 - 9
tid digit, mod digit: ---> 9 - 9

Misrecognized 0 out of 44 utterances.
Accuracy = 100.00%
```

6 HMM Likelihood and Recognition

6.1 Forward Algorithm

We want to calculate the likelihood of the sequence of observations X given the model $\Theta = \{\Pi, \alpha, \Phi\}$. After calculating all the forward probabilities, the probability of generating the entire observation sequence by the HMM is the sum of the forward probabilities for the last time-step for all possible final states. In the end we convert the forward probs into log domain and we calculate the log likelihood of every utterance for every HMM model. By choosing the model that produces the highest likelihood we get the classification.

Figure 3: logalpha

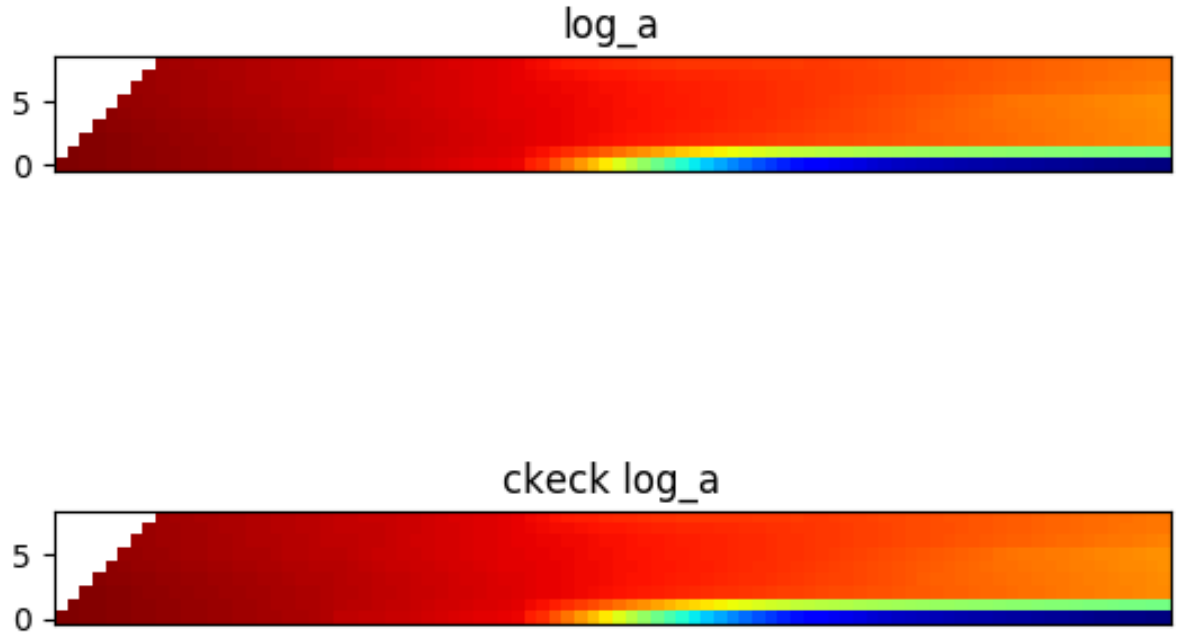
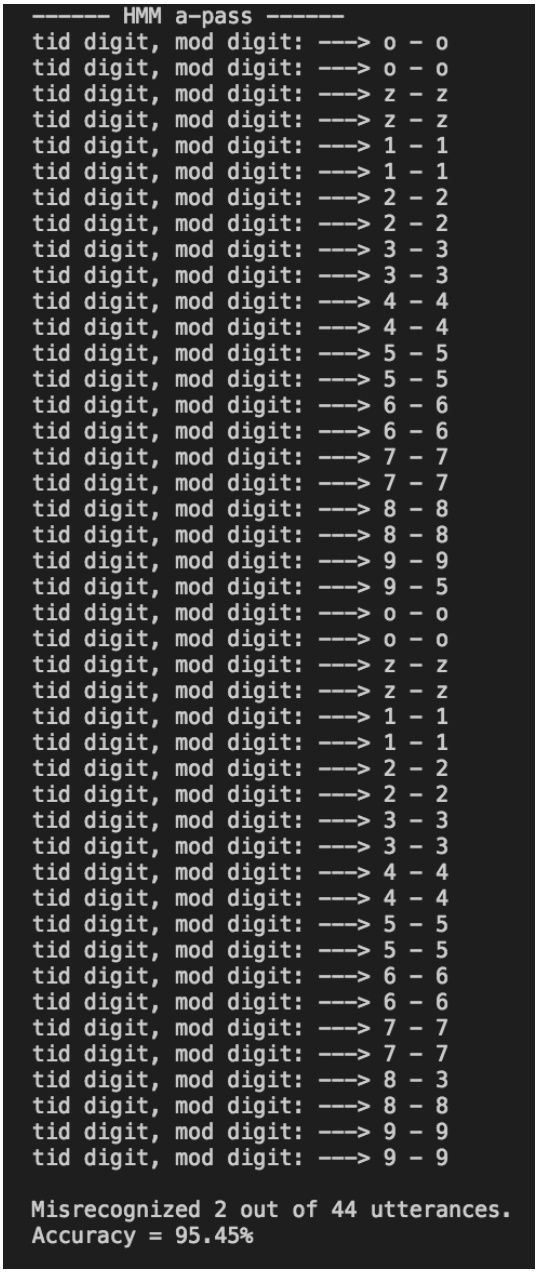


Figure 4: HMM a-pass



Furthermore, we perform a recognition by using the HMMs distributions as if they are GMMs, by setting their weights all equal. In this way, we assume that the transition from any state to any other state is equally probable. Thus we view all possible state sequences as equally probable.

Figure 5: HMM as GMM

```
------ HMM as GMM ------
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 9 - 9
tid digit, mod digit: ---> 9 - 9
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> o - o
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> z - z
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 1 - 1
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 2 - 2
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 3 - 3
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 4 - 4
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 5 - 5
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 6 - 6
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 7 - 7
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 8 - 8
tid digit, mod digit: ---> 9 - 9
tid digit, mod digit: ---> 9 - 9

Misrecognized 0 out of 44 utterances.
Accuracy = 100.00%
```

6.2 Viterbi Approximation

In this task we use the Viterbi approximation in order to calculate the log likelihood of the observation sequence X given an HMM model and the best sequence of states. At every time-step, the log-V gives the probability of the most probable state sequence. Having calculated the log-V for all possible states and for all the time-steps of an observation sequence, we are able to retrieve the likelihood of the most probable path (the highest log-V of the last time-step) and also the state in which the most probable path ends.

With this information we are able to trace back the most probable path until the first state. *Here we can assume that the digit has only one phoneme $(1 + 2) * 3 = 9$ states.

Comparing the results between the Viterbi approximation and forward algorithm, we get the same results. When calculating the Viterbi approcimation, we choose only the most probable state of every time-step in order to calculate the next time-steps probabilities, whereas in the forward algorithm, we combine all possible states for every time-step. This results in different ultimate likelihoods that are, in a way, scaled. But the ranking final probability of every model compared to all the others should remain the same.

Figure 6: HMM logalpha and Best Path

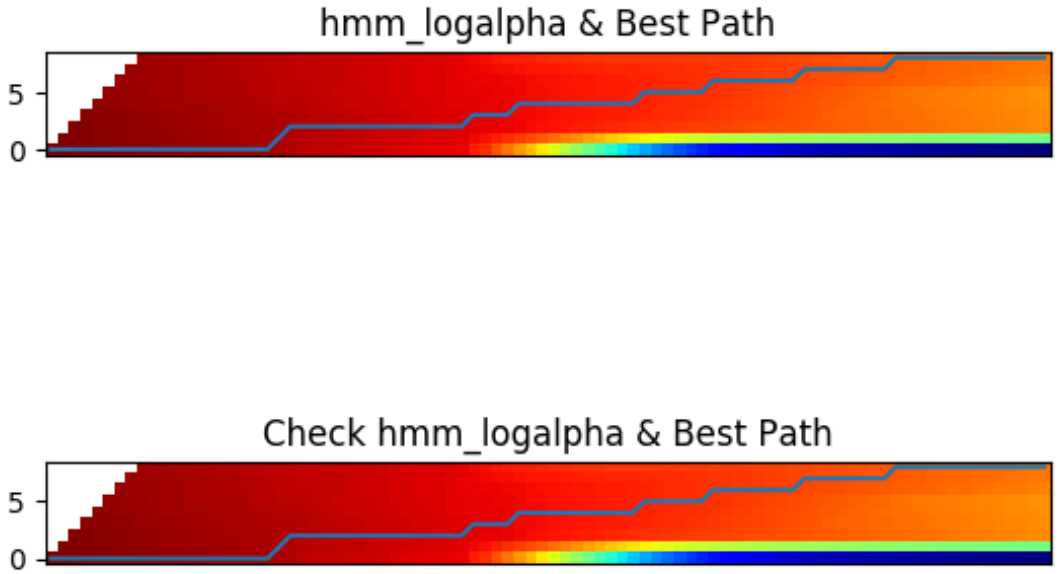


Figure 7: Viterbi approximation