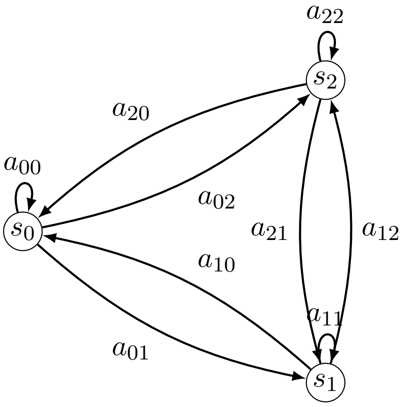
KTH Computer Science
and Communication

Concatenating HMMs

In speech recognition we define hidden Markov models in a hierarchical way. The models are usually defined at the phonemic level, then they are combined to form word models and, finally, the word models are combined to form models of each utterance for training, or using a grammar for recognition. In real recognizers this is usually implemented with specific data structures. However, if we want to stick to the simple definitions of the models (based on matrices or two dimensional arrays) in order to keep the functions simple, we need to understand what this implies in terms of state numbering and transition matrices. This document tries to clarify this in the simple case when the combination of models is restricted to concatenation.

1 Introduction

What we describe below is valid in general, but we will make a concrete example of three state hidden Markov models (HMMs) that are used in practice in speech recognition to model each phoneme in the language. In the general definition this model is depicted by the figure below, where we start indices from 0 to simplify comparison with the Python implementation:



The model consists of the following probability distributions:

$$\Pi = \begin{bmatrix}\pi_0 & \pi_1 & \pi_2\end{bmatrix}, \qquad \text{a priori probability } \pi_i \text{ of state } i$$

$$A = \begin{bmatrix}a_{00} & a_{01} & a_{02}\\ a_{10} & a_{11} & a_{12}\\ a_{20} & a_{21} & a_{22}\end{bmatrix}, \qquad \text{transition probability } a_{ij} \text{ from state i to state j}$$

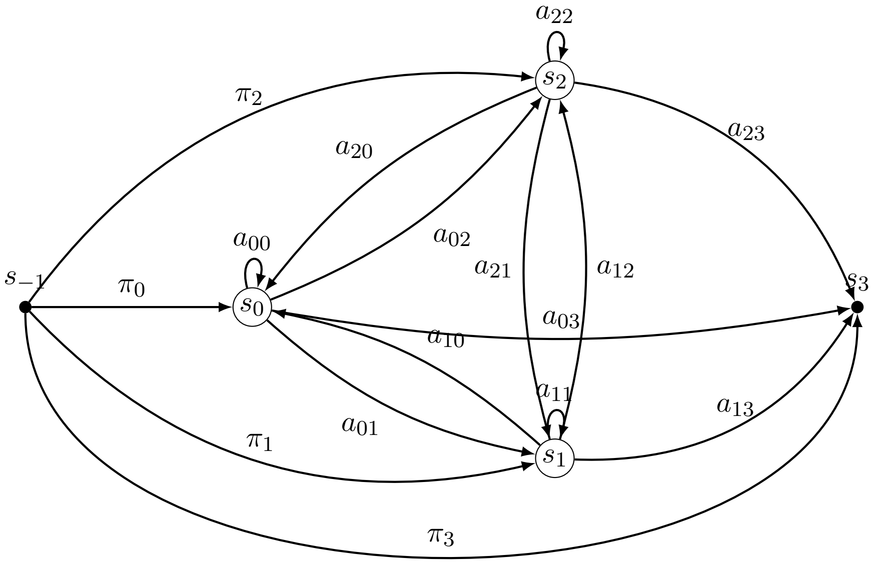
$$\phi_i(x_n) \qquad \text{probability of emitting } x_n \text{ from state } i$$

The emitting probabilities can have different forms, but in the labs they are defined by Gaussian distributions with diagonal covariance matrices, and therefore consist of two 3 × 13 matrices where each row corresponds to a state and contains the means and variances for the 13 features in the observation vectors.

If we want to concatenate this model with another, the first problem we observe is that there is no probability defined for leaving the model. In order to be able to define this, we need to add an output state (s_3 in this case) to the model. This state is said to be *non emitting* because there is no emission probability associated with it. It is only a placeholder for the destination states in the next model. The emission probability distributions will therefore stay the same, but the a priori state probability and the transition probabilities will be as following:

$$\begin{aligned}
\Pi &= \begin{bmatrix}\pi_0 & \pi_1 & \pi_2 & \pi_3\end{bmatrix}, && \text{a priori probability of state } i \\
A &= \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix}, && \text{transition probability from state i to state j}
\end{aligned}$$

The last row in A indicates that reaching s_3 corresponds to exiting the current model. There is therefore no way we can go back to the previous states, unless the combination of models allows that. Also we have now some non-zero probability of exiting the model expressed by a_{i3} for each emitting state s_0, s_1, and s_3. We even have a possibly non-zero probability π_3 of skipping the model altogether without consuming any time step (or observation vector). We can illustrate this model with the following figure, where we have also introduced another placeholder state s_{-1} to include the a priori probabilities of the states in the graph. Non-emitting states are displayed as dots, emitting states as circles:



Note that we could combine both the a priori probability of the states and the transition probabilities into the same matrix, in the assumption that we will always start in s_{-1}[1]:

$$T = \begin{bmatrix} 0 & \pi_0 & \pi_1 & \pi_2 & \pi_3 \\ 0 & a_{00} & a_{01} & a_{02} & a_{03} \\ 0 & a_{10} & a_{11} & a_{12} & a_{13} \\ 0 & a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

[1] This is the approach taken in HTK (`http://htk.eng.cam.ac.uk/`), for example

However, we prefer to keep a priori probabilities and transition probabilities separated to be more consistent with the theoretical definitions. This also means that the state s_{-1} is only used here for illustration and is not defined explicitly in the models.

2 Concatenating two HMMs (`concatTwoHMMs`)

If we want to concatenate two models in the form above, we could simply use the final non-emitting state of the first model as if it was the s_{-1} state in the second model, and therefore, connect this state to any state of the second model using the a priori probabilities of the second model.

We assign different symbols to the a priori and transition probabilities in the two models: π_i and a_{ij} for the first model, and ρ_i and b_{ij} for the second model:

$$\Pi = \begin{bmatrix}\pi_0 & \pi_1 & \pi_2 & \pi_3\end{bmatrix}, \qquad P = \begin{bmatrix}\rho_0 & \rho_1 & \rho_2 & \rho_3\end{bmatrix},$$

$$A = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} b_{00} & b_{01} & b_{02} & b_{03} \\ b_{10} & b_{11} & b_{12} & b_{13} \\ b_{20} & b_{21} & b_{22} & b_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

With the above method, the resulting model has a number of states equal to the sum of the total states in the two original models and prior probability and transition matrix is as follows:

$$\Pi_{\text{concat}} = \begin{bmatrix}\pi_0 & \pi_1 & \pi_2 & \pi_3 & 0 & 0 & 0 & 0\end{bmatrix},$$

$$A_{\text{concat}} = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} & 0 & 0 & 0 & 0 \\ a_{10} & a_{11} & a_{12} & a_{13} & 0 & 0 & 0 & 0 \\ a_{20} & a_{21} & a_{22} & a_{23} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_0 & \rho_1 & \rho_2 & \rho_3 \\ 0 & 0 & 0 & 0 & b_{00} & b_{01} & b_{02} & b_{03} \\ 0 & 0 & 0 & 0 & b_{10} & b_{11} & b_{12} & b_{13} \\ 0 & 0 & 0 & 0 & b_{20} & b_{21} & b_{22} & b_{23} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that, if we include the prior in the transition matrix this is a block matrix with each block corresponding to a single model. The only exception is the probability a_{33} of staying in the last non-emitting state of the first model. It used to be equal to 1, but now the probability mass has been spread to the a priori probabilities of the states in the second model.

Although this is the most simple way to combine HMMs, it makes the states of the resulting model a mixture of emitting and non-emitting states. All the functions for inference and learning need to be implemented to take this into account.

In the implementations in the labs, instead, we are assuming that the only non-emitting state in the model is the last, and we can simplify the implementation of the functions. In order to achieve this, we need to eliminate the non-emitting state between the two models (s_3 in the example above).

In order to do this, we need to make sure that each transition into the final state in the first model can reach any emitting or final state in the second model. The actual transition probability from state s_i in the first model to state s_j in the second model will be the product of the probability π_{N-1} or $a_{i(N-1)}$ (π_3, a_{i3} in the example) of leaving the first model from state i and the probability ρ_j of entering the second model in state j.

The prior vector and the transition matrix of the concatenation of the two models using this method is:

$$\begin{aligned}
\Pi_{\text{concat}} &= \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3\rho_0 & \pi_3\rho_1 & \pi_3\rho_2 & \pi_3\rho_3 \end{bmatrix}, \\
A_{\text{concat}} &= \begin{bmatrix}
a_{00} & a_{01} & a_{02} & a_{03}\rho_0 & a_{03}\rho_1 & a_{03}\rho_2 & a_{03}\rho_3 \\
a_{10} & a_{11} & a_{12} & a_{13}\rho_0 & a_{13}\rho_1 & a_{13}\rho_2 & a_{13}\rho_3 \\
a_{20} & a_{21} & a_{22} & a_{23}\rho_0 & a_{23}\rho_1 & a_{23}\rho_2 & a_{23}\rho_3 \\
0 & 0 & 0 & b_{00} & b_{01} & b_{02} & b_{03} \\
0 & 0 & 0 & b_{10} & b_{11} & b_{12} & b_{13} \\
0 & 0 & 0 & b_{20} & b_{21} & b_{22} & b_{23} \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\end{aligned}$$

If we observe for example the colored part of the state prior, in those cases, we are skipping completely the first model (with probability π_3) and we are starting the second model in any of its states with the corresponding probabilities ρ_j. Similarly, in the first row of the transition matrix, we are exiting the first model from state s_0 (with probability a_{03}) and we are entering the second from state s_j, again, with probability ρ_j.

In this case, all the states in the resulting model but the last are emitting, and the number of states is the sum of the emitting states in the original models plus one.

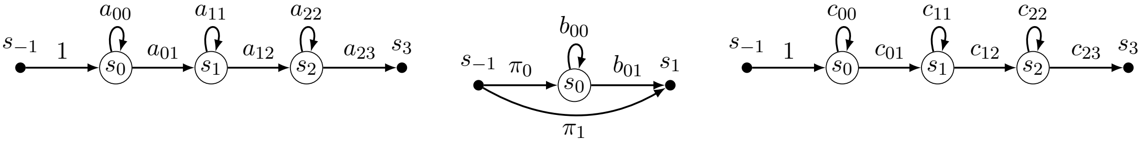
Because all the emitting states in the resulting model are still contiguous, the matrices describing the emission probability distributions can simply be concatenated in the state dimension (which is the row index in our definition).

In Lab 2, the function implementing this concatenation is called `concatTwoHMMs()`. If we want to concatenate several HMMs, we can repeat the simple function iteratively, as is done in `concatHMMs()`.

3 A practical example

The above discussion is valid for generic HMMs where we can reach any emitting state from any other and we can enter and exit the model from any emitting state. Here we give a practical example with left-to-right models that are usually employed in speech recognition. We also include a special model with only one state and skip connections that is used for optional short pauses sp between words (this will be used in Lab 3).

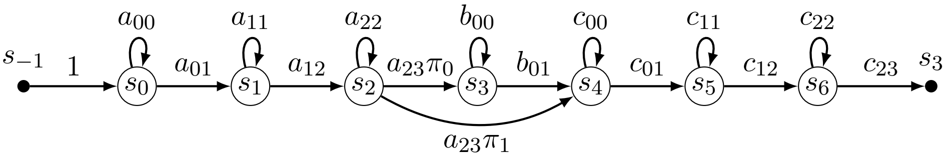
The models are illustrated by the graphs and definitions below:



$$\Pi_a = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \qquad \Pi_c = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} a_{00} & a_{01} & 0 & 0 \\ 0 & a_{11} & a_{12} & 0 \\ 0 & 0 & a_{22} & a_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{aligned} \Pi_b &= \begin{bmatrix} \pi_0 & \pi_1 \end{bmatrix} \\ B &= \begin{bmatrix} b_{00} & b_{01} \\ 0 & 1 \end{bmatrix} \end{aligned} \quad C = \begin{bmatrix} c_{00} & c_{01} & 0 & 0 \\ 0 & c_{11} & c_{12} & 0 \\ 0 & 0 & c_{22} & c_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The resulting model is:



$$\begin{aligned} \Pi_{\text{concat}} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \\ A_{\text{concat}} &= \begin{bmatrix} a_{00} & a_{01} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{22} & a_{23}\pi_0 & a_{23}\pi_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_{00} & b_{01} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{00} & c_{01} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{11} & c_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c_{22} & c_{23} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$