Quiz 4: Markov Models - Hidden Markov Models

Introduction to Supervised Learning

Email address *

mikesymmonds@gmail.com

Please enter your name: *

Michael Symmonds

Generative Classifier

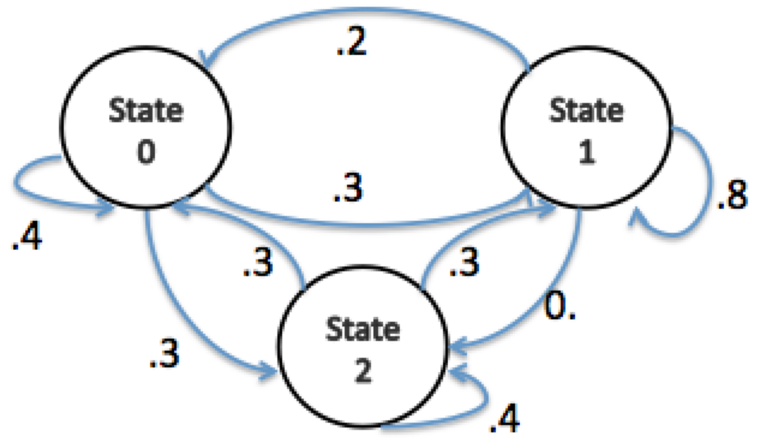
We want to create a generative classifier for sequential data as explained in the lecture. To that end, we learn two models: Model 0 (parameterized by theta_0) on the samples associated to the target 0 and Model 1 (parameterized by theta_1) on the samples associated to the target 1. Calculate the probability of getting a positive target for the sequence X_1 ... X_T using the following probabilities: 1 point

$$\left.\begin{array}{l} p_{\theta_0}(X^1,\ldots,X^T|Y=0)=10^{-1} \\ p_{\theta_1}(X^1,\ldots,X^T|Y=1)=10^{-2} \\ p(Y=1)=0.7 \end{array}\right\} \quad p(Y=1|X^1,\ldots,X^T) \ ?$$

7/37 = 0.189

Markov Models

After fitting a Markov Model on sequences of discrete data (with 3 possible states for each sample). We end up with the following graph.



What is the number of parameters of a Markov Model with 3 states? 1 point

- ○ 6
- ○ 9
- ◉ 12

What is the probability of transitioning from state 2 to state 0? 1 point

- ○ 0.2
- ◉ 0.3
- ○ 0.4

If we are at state 1, what is the most likely next state? 1 point

- ○ State 0
- ◉ State 1
- ○ State 2

Explain why it is impossible to have this sequence of states in the training data: 1 2 0 0 2. 1 point

The transition from state 1 to state 2 has zero probability and as such, the sequence described above is impossible. The first jump is not possible.

Hidden Markov Models

We want to use an HMM model to fit discrete observations taking values in {0, 1, 2}.

What would be the number of parameters of the HMM model if we use M hidden states? (as a function of M) 1 point

M+M*M+M*3 (Here, V=3 since we have 3 possible observations)

After fitting an HMM with 3 hidden states on the previous observations, we end up with the following parameters.

$$\pi = \begin{pmatrix} 0.7 \\ 0.2 \\ 0.1 \end{pmatrix} \qquad Q = \begin{pmatrix} 0.7 & 0.1 & 0.2 \\ 0.3 & 0.6 & 0.1 \\ 0.1 & 0.2 & 0.7 \end{pmatrix} \qquad O = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \end{pmatrix}$$

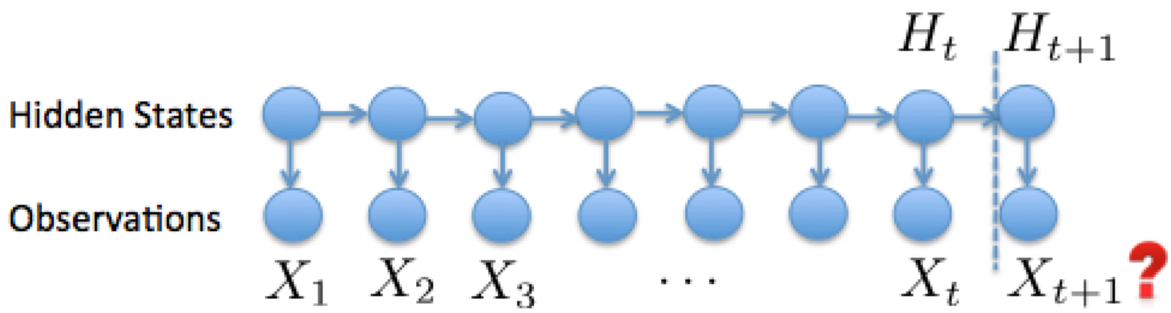
What is the probability of transitioning from the hidden state 0 to the hidden state 1 ? 1 point

- ◯ 0.7
- ◉ 0.1
- ◯ 0.2

What is the distribution of the discrete observations conditioned on the hidden state 0 ? 1 point

- ◉ [0.2, 0.3, 0.5]
- ◯ [0.2, 0.8, 0.1]
- ◯ [0.7, 0.2, 0.1]

We want to predict the next observation X(t+1) based on the observations X(1) .. X(t)



Using the following filtering probabilities and the previous fitted parameters, what is the most likely next observation X(t+1) ? Justify yout answer. 2 points

$$[p(H_t = h | X_1 = x_1, \ldots, X_t = x_t)]_{h\in\{0,1,2\}} = \begin{pmatrix} 0.7 \\ 0.2 \\ 0.1 \end{pmatrix}$$

To find the most likely next observation X(t+1), we first need to find the probability distribution of the next hidden state, i.e. p(Ht+1=h|X). This probability is given by the dot product of Q' and the filtering probabilities given above. We need to transpose Q so that each row of the resulting matrix represents the probability of moving into the hidden state represented by the row. This also ensures the probabilities add up to one. Hence, multiplying Q'*p(Ht=h|X) gives, p(Ht+1|X)= [0.56 , 0.21 , 0.23] . We now use this vector to find the distribution of the next observations. To do this, we use the observation matrix above. We calculate the dot product of O' and p(Ht+1|X), again taking the transpose of O. This gives, p(Xt+1|Xt) = O'*p(Ht+1|X) = [0.303 , 0.35 , 0.347] . Again our probabilities sum to one. Looking at this vector, the most likely next observation is 1.

Programming Session

Did you understand the problem?

- ◉ Yes
- ◯ No

Google