# Assignment #8 – Machine Learning – Professor Haugh

## **Jaime Gacitua**

## **jg3499**

## **Due Monday May 2, 2016**

## Question 1

1. Computing is solving the likelihood problem. This can be solved by using the relationship

We have , and

To get there we have to iterate over the.

For,

Moving to,

Finally, on,

Finally, using the relationship presented before,

1. To find we need to solve the smoothing problem – inferring the past. To get this value, we have to solve for the following relationship:

We already calculated. We have to do the backward recursion to get the:

For,

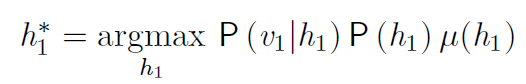
For,

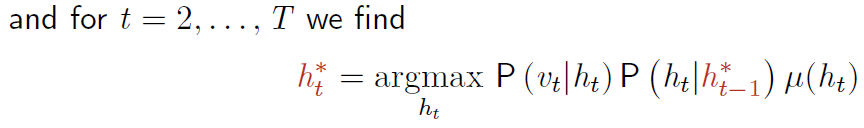
For,

Finally, using the above formula,

1. The most likely path problem can be solved using the Viterbi Algorithm,

Once we have solved all the way to we backtrack to obtain the most likely hidden path. We get





Solving the sequence, we get the most probable hidden state sequence is

The calculations can be found below.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | t=3 |  |  |  | |  | h\_(t-1) |  |  | | h\_(t) | 1 | 2 | 3 | | 1 | 0.350 | 0.000 | 0.000 | | 2 | 0.120 | 0.240 | 0.000 | | **3** | 0.160 | **0.320** | 0.800 | | max | 0.3500 | 0.3200 | 0.8000 | |  |  |  |  | | h3 = |  | 3 |  | | |  |  |  |  | | --- | --- | --- | --- | | t=2 |  |  |  | |  | h\_(t-1) |  |  | | h\_(t) | 1 | 2 | 3 | | 1 | 0.053 | 0.000 | 0.000 | | **2** | **0.058** | 0.115 | 0.000 | | 3 | 0.032 | 0.064 | 0.160 | | max | 0.058 | 0.115 | 0.160 | |  |  |  |  | | h2 = | 2 |  |  | |
| |  |  | | --- | --- | | t=1 |  | |  | a | | h\_(t) | 1 | | **1** | **0.036** | | 2 | 0.005 | | 3 | 0.000 | | max | 0.036 | |  |  | | h1 = | 1 | |  |

## Question 2

1. First we use the definition of conditional probability, ,

Again, we use the same definition, but in the form,

By the structure of the HMM model, is independent from, therefore,

On the denominator, we use the law of total probabilities, conditioning over,

On the denominator, we use the definition of conditional probability, as, yielding,

By the structure of the HMM model, is independent from, therefore, we finally arrive to

With this expression, we see that

Represents the transmission matrix. Known.

Represents the solution to the filtering problem:

Assuming we have already solved for,

Finally,

1. Re-writing the smoothing problem by using total probabilities over ,

Using the definition of conditional probability,

Because of the structure of the HMM, is independent of, therefore,

If we define, then we can rewrite the last relationship as

We can then solve iteratively for, working backwards.

In addition, we must be clear on how to define the gamma of period, which is the same as the solution to the filtering problem,

## Question 3

Suppose the state space of the hidden sequence is of cardinality, and the visible state space has cardinality.

Suppose we start with an initialization vector , the transition matrix, and emission matrix to uniformly constant values, say, , and, .J

Computing the pairwise marginal and marginal, we use the expression derived in lecture,

We know and.

Solving for,

Generalizing,

Slving for,

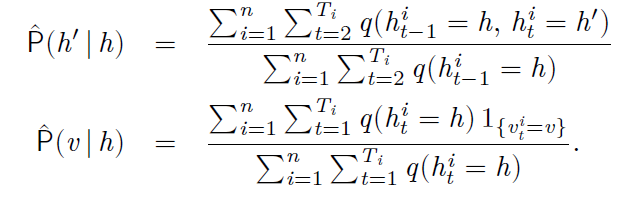
Generalizing,

Using these solutions, we go back to calculate the pairwise marginal,

Normalizing,

We can get the marginal through the pairwise marginal,

When we update the parameters in the M-step, we have to apply the following equations,



Therefore,

If we start with uniformly constant values,

* The update of the matrix will be meaningless, because all of its elements will stay unchanged as.
* The update of the matrix will be meaningless, because all of its elements will have the same value, for all time periods and hidden states, with a value equal to,

## Question 4

1. When we have Gaussian emissions, with mean and standard deviation depending on the hidden state k,

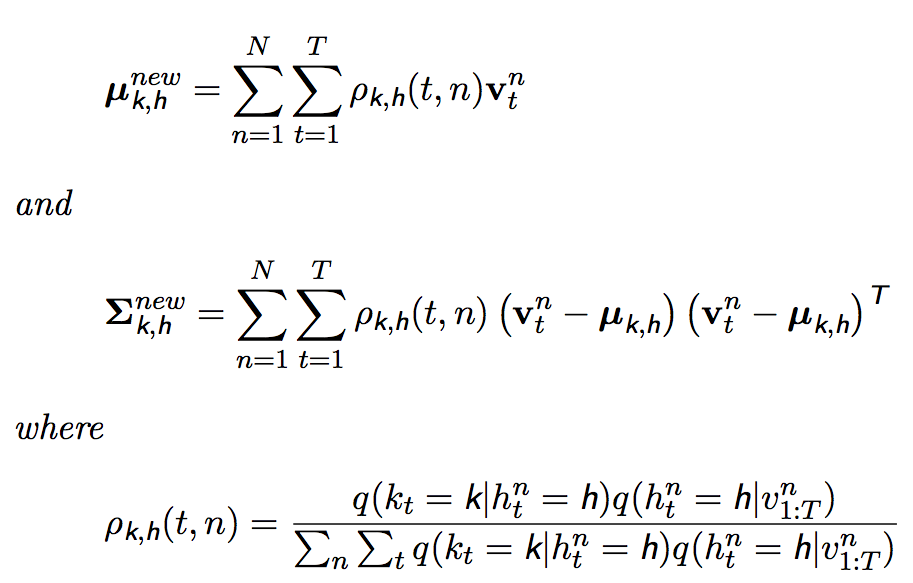
Calculating the derivative with respect to, and equaling to zero,

gma 𝑚𝑎 ing tandard deviation depending on the discrete state k,Calculating the derivative with respect to, and equaling to zero,

1. There is no difficulty when extending the EM algorithm to the case where the emission distribution is a mixture of Gaussians.

We have to add a new variable to the emission distribution, indexed as. This will be a considered as a latent variable and can be introduced in the following way into the HMM model:

The EM algorithm continues in a straightforward way, carrying over this extension, to reach the following result:



## Bibliography

Barber, D., 2015, Bayesian Reasoning and Machine Learning.

Petersen et. Al., 2012, The Matrix Cookbook

James et. Al., 2015, Introduction to Statistical Learning

Hastie et. Al., 2001, The Elements of Statistical Learning