Homework 5

Anusha Ramamurthy

October 16, 2016

1. **Hidden Markov Models**

   One application of HMMs is cryptanalysis. Imagine a simple substitution code, in which every character of the English alphabet is substituted by another character, thus Banana could be translated into Ucpcpc. In order to decode the encrypted word, a bigram model could be used.

   a.) Assuming you do not know the code beforehand – i.e., your task is to decrypt a message – describe what the HMM would look like, in order to decode such words. How many and which states will the HMM have, and what do the emissions look like?

   Answer:

   The HMM would consist of Initial State Distribution given by

   $$w_i = P(Q_0 = i)$$

   where w_i is the probability of the state i at time zero.This can have any prbability distribution, as research has shown that for very large t, the initial probabilities don't matter.

   Transition Probabilities as

   $$p_{ij} = P(Q_{t+1} = j|Q_t = i)$$

   where the probability of transition from state i to j is the probability of the new state being j given the previous state was i. Assuming the message is an encoding of the english language, we could look for bigrams that frequently precede another bigram and assign transition probabilities in a corresponding manner by looking at a very large corpus.

   Emission Probabilities as

   $$e_i(a) = P(O_t = a|Q_t = i)$$

   where the emission probability of an observation of a bigram a at time i is the probability of the observation being one of the bigrams given that the state was i.

   b.) Make a suggestion for how the emission probabilities should be distributed for a well-trained HMM.
   Answer:
   We can define the emission probabilities to be the observations at a given state. The emission probabilities will be distributed similar to the distribution of bigrams for an english corpa, with the most frequent bigrams in english being the bigrams with the higher emission probabilities.

   c.) Would it make sense to convert the HMM to a trigram model instead of a bigram model? What are the advantages, what are the disadvantages?

Answer:

i. I would argue that it depends on the context and the data to which we wish to use HMM on.

ii. Advantages: As Triagrams are a special case of bigrams, with the trigram model we might achieve better prediction and accuracy.

iii. Disadvantages: Trigrams may cause the HMM to over fit to the data, for we now expect three words to often be associated together.

iv. Suggession: I would advocate using the bias-variance approach, where we check if the error on training set and testing set is high. Then the model is over fitting to data set and shows high bias. In such a context I would suggest using a bigram model. If there is a large difference in the error between the test and train sets, then there is high variance and we can increase the complexity of the model.

2. Forward/Backward Algorithm

Consider the example of the HMM for POS tagging the sentence a myth is a female moth. Calculate the following forward accumulators given the following probabilities:

Compute the forward (α) and backward (β) probability for the given time steps and states:

Answer

(a) $\alpha_4(NN) = 0.0003$

(b) $\alpha_3(VB) = 0.018$

(c) $\alpha_1(DT) = 0.3825$

(d) $\beta_4(NN) = 0.01$

(e) $\beta_2(NN) = 0.006$

Answers are computed using the program attached. Please note that the program was written in collaboration with Anup Bharadwaj.