

Hidden Markov models

Nikhil Jangamreddy (2018csm1011), Aakash Saini (2018csm1001)

May 7, 2019

1 Loading and Writing the model files:

To implement the model we have written class **HMM** which denotes the Hidden Markov Model. For this class transition and emission probabilities are used as the parameters which are specified in the files like `models/two_english.trans` and `models/two_english.emit`. In `.emit` file, each row is having symbol which denotes the start state.

Function `HMM.load` is used to load the parameters from the model files in which some of the transitions may be omitted. Such parameters maintains value 0. If conditional probabilities are not specified in model files then constructor will initialize them randomly. Inverse method **`HMM.dump`** takes basename of output file model's parameter are written to the `.trans` and `.emit` files.

2 HMM Supervised learning :

For supervised learning method **`HMM.learn_supervised`** is defines which takes which takes observation's list with know state sequences from the file *browntags.obs*. This method also estimate the parameters of HMM using the maximum likelihood estimator.

3 Generating random observations :

We have defined method **`HMM.generate()`** which generates the sampling from the HMM. This function takes an integer *n* as input parameter and returns a random observation having length *n*.

```

hmm.py
DET NOUN . NOUN VERB DET NOUN NOUN ADP ADV DET NOUN ADP ADJ
what exposures , food cut the office plant of altogether the market of internal
. . NOUN . . CONJ DET NOUN ADP DET
, , forces , , and an taxpayers of the
DET ADJ NOUN . ADV . CONJ NOUN
the small area , well . and continents
PRON . DET VERB PRON VERB CONJ VERB NOUN PRT VERB
then ' ' a been he pass and landed voice today'll apportioned
PRON VERB ADP CONJ NUM
he is in and 100
ADP PRON VERB
around he refused
ADP NOUN . VERB ADV PRT ADV . ADV ADP DET
in fear , could not to lengthwise , partially of her
DET NOUN
the cancer
VERB NOUN VERB VERB
wished performers held sent
NOUN NOUN VERB NOUN . ADP DET ADV DET NOUN . PRON VERB
town entry might government , by the however the troubles , who been
NOUN VERB ADV
flrms gave just
CONJ
and
. CONJ NOUN NOUN . NOUN . NOUN ADP DET VERB DET NUM ADJ
. or kayabashi igor , intensity . print through this is some 1960 present
NOUN
variety
DET NOUN PRON . NUM NOUN VERB ADP NOUN . ADP DET NOUN
the sunday me ' ' 1961 pressure accepted for system . for the part
PRON VERB ADP DET NOUN CONJ NUM NOUN VERB VERB ADV VERB VERB
ll laundering for the research but two joy was cannot ago knowing tried
ADJ NOUN . NOUN ADP PRT
american af . peer-group of toot
PRON VERB ADV VERB ADP DET NOUN CONJ
he received softly are of the pictures and
. VERB VERB DET NOUN ADP NOUN
. don't filled that teats of bottle
ADV PRT . . CONJ NOUN ADP DET NOUN . .
strongly i'n . , and mana for some laos , ,

```

Figure 1: Random 20 Observations

4 Finding best State sequence using Viterbi algorithm:

For finding the best State sequence using Viterbi Algorithm, we have defined a method named **HMM.viterbi()** for implementing the Viterbi algorithm. This method assigns the observation's state sequence with Viterbi state sequence.

```

PRON VERB DET NOUN .
i shot the elephant .
PRON VERB DET NOUN ADP DET NOUN .
he took my shot at the elephant .
NOUN VERB ADP DET NOUN .
flies waited at the window .
DET NOUN VERB DET NOUN .
the pilot flies the plane .
DET VERB DET ADJ NOUN .
this is a light blanket .
PRON VERB DET NOUN PRT .
she turned the light off .
DET NOUN NOUN DET NOUN .
the lanterns light our path .
VERB PRON VERB PRON .
did you train her ?
DET NOUN VERB VERB ADV .
the train is arriving now .
PRON NOUN DET NOUN .
they book the ticket .
PRON VERB DET NOUN .
i love this book !

```

Figure 2: Output file in Viterbi algorithm

5 Implementing Forward and Backward Algorithms :

Forward and Backward algorithms are implemented in the functions **forward** and **backward()** respectively, both these methods take output sequence as input.

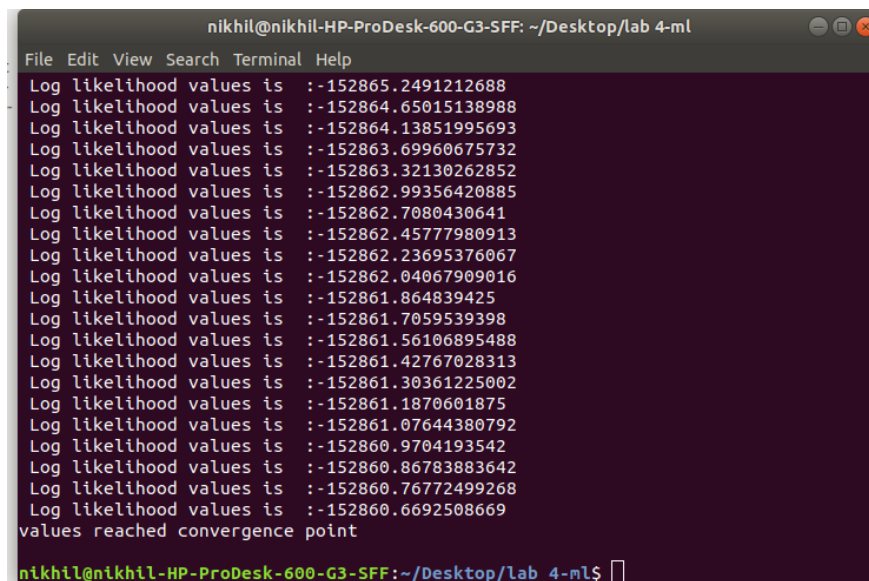
We have also defined the methods **forward_probability()** and **backward_probability()** for calculating the probability for a given output sequence using the values returned from *forward* and *backward* functions.

The probability estimated by the *forward_probability* function from the file *ambiguous_sents.obs* is written to the file *ambiguous_sents.forwardprob*.

The probability estimated by the *backward_probability* function from the file *ambiguous_sents* is written to the file *ambiguous_sents.backwardprob*.

6 Supervised and Unsupervised Learning with Baum Welch(EM) algorithm:

Function **learn_unsupervised()** is defined for estimating the parameter of model from the beginning of current model and log-likelihood for the trained model. This method takes a list of observations for which state sequence is unknown, convergence threshold, booleans for specification of transition and emission parameters, and the number of random restarts. Baum Welch EM algorithm is used by this algorithm by drawing upon the backward and forward algorithm.



```
nikhil@nikhil-HP-ProDesk-600-G3-SFF: ~/Desktop/lab 4-ml
File Edit View Search Terminal Help
Log likelihood values is :-152865.2491212688
Log likelihood values is :-152864.65015138988
Log likelihood values is :-152864.13851995693
Log likelihood values is :-152863.69960675732
Log likelihood values is :-152863.32130262852
Log likelihood values is :-152862.99356420885
Log likelihood values is :-152862.7080430641
Log likelihood values is :-152862.45777980913
Log likelihood values is :-152862.23695376067
Log likelihood values is :-152862.04067909016
Log likelihood values is :-152861.864839425
Log likelihood values is :-152861.7059539398
Log likelihood values is :-152861.56106895488
Log likelihood values is :-152861.42767028313
Log likelihood values is :-152861.30361225002
Log likelihood values is :-152861.1870601875
Log likelihood values is :-152861.07644380792
Log likelihood values is :-152860.9704193542
Log likelihood values is :-152860.86783883642
Log likelihood values is :-152860.76772499268
Log likelihood values is :-152860.6692508669
values reached convergence point
nikhil@nikhil-HP-ProDesk-600-G3-SFF:~/Desktop/lab 4-ml$
```

Output of final log likelihood