# CS 688: Graphical Models - Spring 2016

## Assignment 2: Part B Report

Shamya Karumbaiah (Collaborated with Rafael Lizarralde)

**1)** *Code - Q1_3optimize_plot.m*
The error rate of the learned model averaged over all predicted characters in all test words is 3.2377 (1046 correct predictions out of 1081 charcters). I used limited memory BFGS from minFunc in Matlab. The objective function value or the maximum log likelihood is -9.96086e-05 after 34 iterations. Using an L2 regularization also resulted in a similar error rate.

**2)**
**1.** I would prefer to use MBR over Viterbi when the given training data is more noisy. Since the algorithm tries to maximize the overall prediction in Viterbi, a noisy data (especially the one that has corrupted only a certain parts of the overall character sequence) will have a worse effect in Viterbi than in MBR.
**2.** In general, since our goal is a structured prediction of the complete character sequence, Viterbi may help to get a better overall prediction than MBR. Also, reading a little about max-product algorithm, it feels like its a more stable algorithm for graphs with loops.

**3)** *Code - Q1_3optimize_plot.m*
**3.1** The $10 \times 10$ transition parameters array $W^T$ visualized as a gray-scale image is below. The y axis refers to the left character and the x axis refers to the right character.

**Q3.1 Learnt transistion paramter in original oredering**
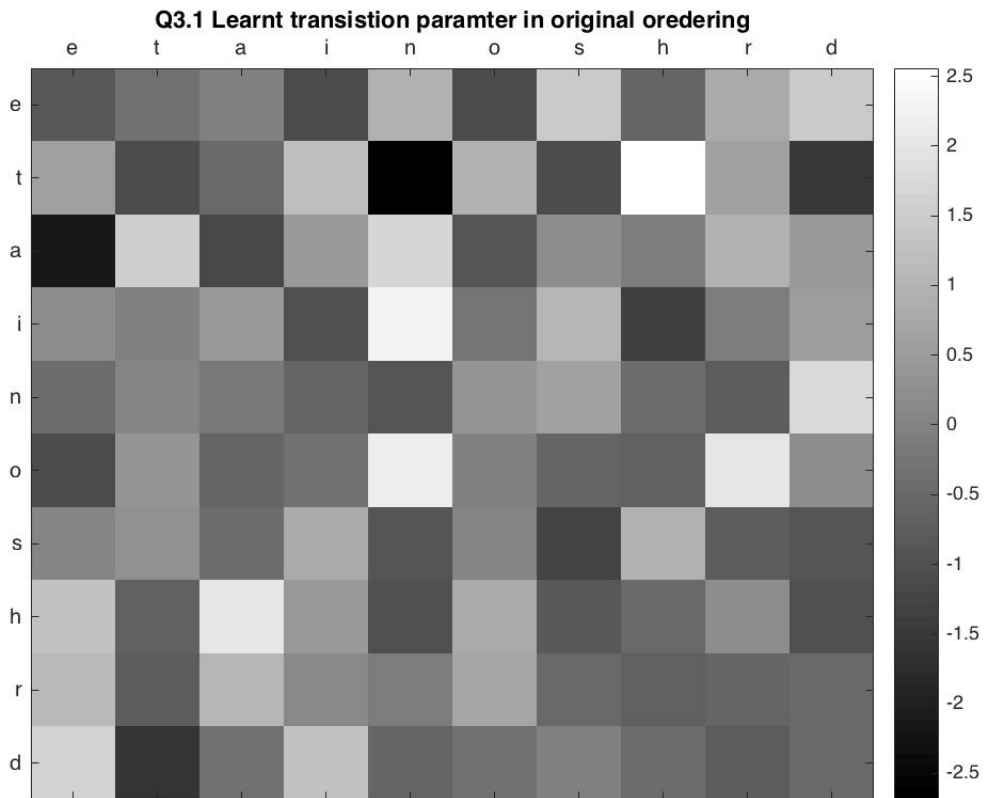


**3.2** Visualization in the order "EAIOTNSHRD" is below. The y axis refers to the left character and the x axis refers to the right character.

**Q3.2 Learnt transistion paramter in new oredering with vowels first**
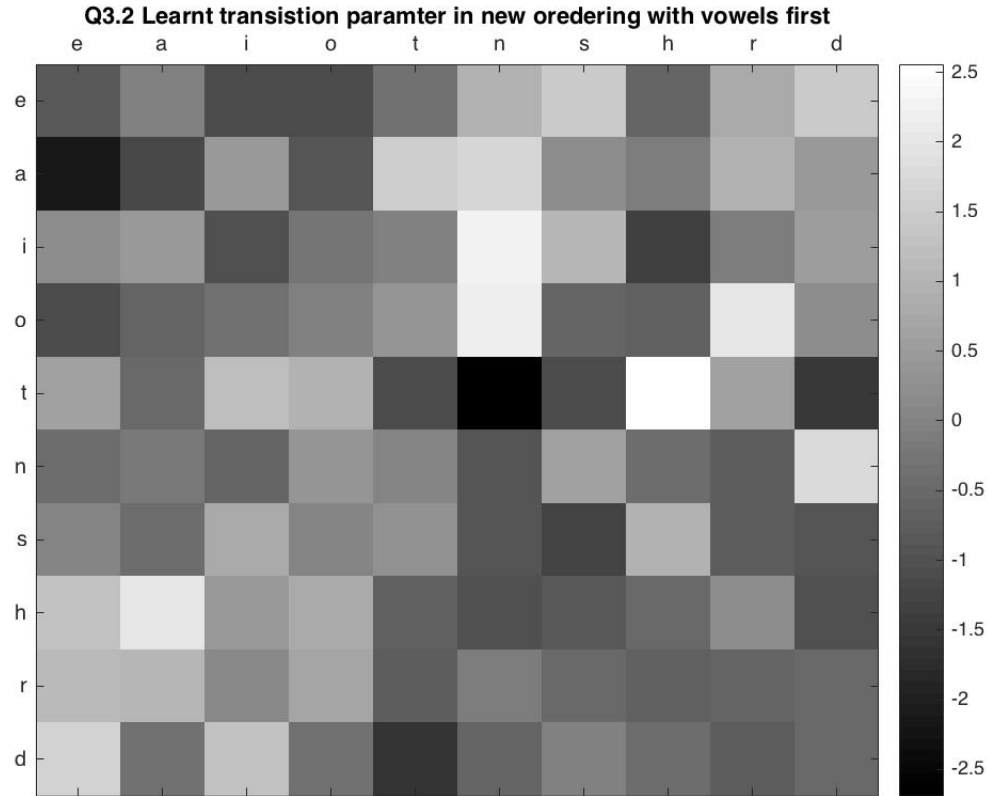


This visualization helped in the observation of the following -

- In general, consonants like to have vowels to their right.

- In most cases, vowels preferred to have consonants to their right though its not as obvious as in the reverse case.

- Few transitions with the largest parameter value are TH, OR, ON, IN, HA, ND, AN. This reasonable because these two pairs occur very often in words. Since the training data had meaningful words, it is explainable that the model learnt to give maximum weights to these transitions.

- In general, N has a overall higher probability to be on the right of a vowel.

- Vowel-vowel pairs and consonant-consonant pairs are very rare.

- Few least possible transition with lowest weight is TN, AE, TD, DT, IH, SS. Its rare to have of words with these pairs.

**4)** *Code - Q4SGD.m*

At the end of the 400th epoch of SGD, the error rate was 0.023127. Below is the plot showing prediction error versus training epoch.

Test error with 400 epochs of stochastic gradient descent
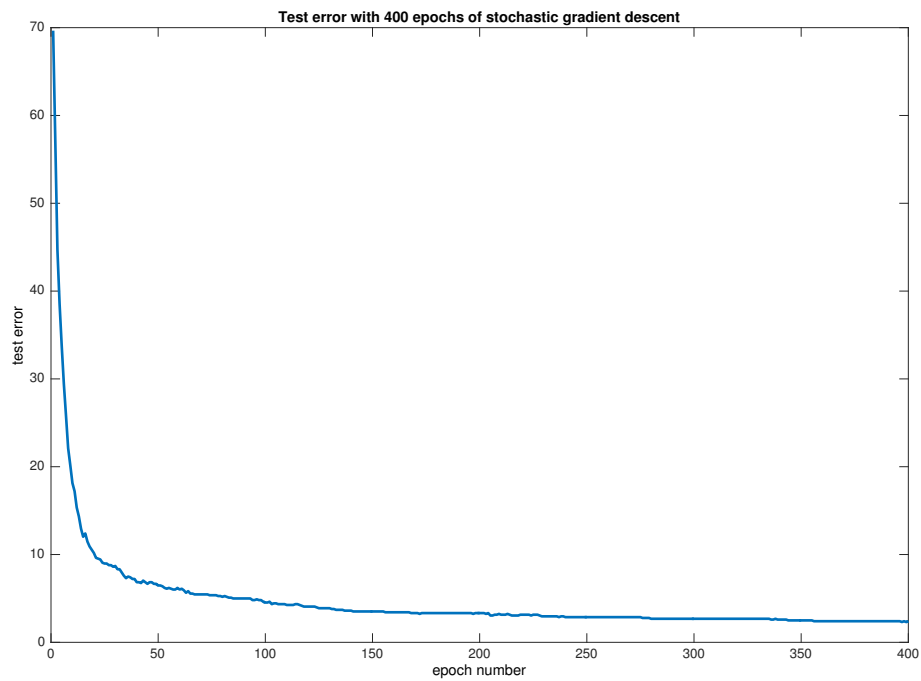
We can observe the wiggle at the end when SGD tries to reach the optima. Below is the plot of the last 300 epochs.



Test error in the last 300 epochs of stochastic gradient descent