Q1.

|  |  |  |
| --- | --- | --- |
| X=1 | X=2 | X=3 |
| .15 | .35 | .5 |

1. =

|  |  |
| --- | --- |
| Y=a | Y=b |
| .6 | .4 |



|  |  |  |  |
| --- | --- | --- | --- |
|  | X=1 | X=2 | X=3 |
| Y=a | .167 | .333 | .5 |
| Y=b | .125 | .375 | .5 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | X=1 | X=2 | X=3 |
| Y=a | .667 | .571 | .6 |
| Y=b | .333 | .429 | .4 |

1. No.



These are not equal.

Q2.

2. .5

Additionally, for p = .5, so this is a local maximum.

Q3.

c1.

c2.

Q4.

1. T states, corresponding to the number of POS tags

Q5.

1. is a word*;* is a tag

<Mike NN w-1 <s> w0 Mike w1 likes w-1w+1 BOS\_likes t-1 BOS t--2t-1 BOS\_BOS>

<likes VBP w-1 Mike w0 likes w1 cats w-1w+1 Mike\_cats t--1 NN t-2t-1 BOS\_NN)>

<cats NNS w-1 likes w0 cats w1 EOS w-1w+1 likes\_EOS t-1 VBP t-2t-1 NN\_VBP>

Q6.

1. I would build a classifier, where x is a document and y is a language. I would choose the following features:
2. Word unigrams
3. Word bigrams
4. Word trigrams
5. Letters
6. Word length
7. Document size
9. Languages poorly represented in training data
10. Documents in closely related languages
11. Some languages are harder than others to identify word boundaries