

NLP - Fall 2017 - Midterm Exam

This material is ©Anoop Sarkar 2017.

Only students registered for this course are allowed to download this material.

Use of this material for “tutoring” is prohibited.

(1) Language Models

Mrs. Malaprop would like to build a spelling corrector focused on the particular problem of *there* vs *their*. The idea is to build a model that takes a sentence as input, for example:

1. He saw their football in the park
2. He saw their was a football in the park

For each instance of *their* or *there* Mrs. Malaprop wants to predict whether the true spelling should be *their* or *there*. So for sentence (1) the model should predict *their*, and for sentence (2) the model should predict *there*. Note that for the second example the model would correct the spelling mistake in the sentence. Mrs. Malaprop recently took some NLP classes so she wants to use a language model for this task. Given a language model $p(w_1, \dots, w_n)$, return the spelling that gives the highest probability under the language model. So for example for the second sentence we would implement the rule: replace *there* with *their* and vice versa and compare the language model scores:

```
If   p(He saw there was a football in the park) >
      p(He saw their was a football in the park)
Then return(there)
Else  return(their)
```

Mrs. Malaprop decides to use an unigram model: $p(w_1, \dots, w_n) = \prod_{i=1}^n q(w_i)$ where $q(w_i) = \frac{\text{Count}(w_i)}{N}$ and $N = \sum_w \text{Count}(w)$. $\text{Count}(\cdot)$ returns the number of times a word was seen in the corpus and N is the sum of counts for all words in the corpus. Assume $N = 10,000$ and $\text{Count}(\text{there}) = 110$ and $\text{Count}(\text{their}) = 50$. Also assume that for every word w in the vocabulary $\text{Count}(w) > 0$.

- a. What does the Mrs. Malaprop rule return for *He saw their was a football in the park*?

Answer: *there*

- max mark: 1 (all or nothing)

- b. Is the Mrs. Malaprop rule a good solution to the *their* versus *there* problem? Say yes or no and give a short and precise one sentence justification for your answer.

Answer: No. Gives the wrong answer for *He saw their football in the park*.

- max mark: 4
- yes / no: 1 mark
- explanation makes sense: 3 marks

(2) Hidden Markov Models

The probability model $P(t_i | t_{i-2}, t_{i-1})$ is provided below where each t_i is a part of speech tag, e.g.

$P(D | N, V) = \frac{1}{3}$. Also provided is $P(w_i | t_i)$ that a word w_i has a part of speech tag t_i , e.g. $P(\text{flies} | V) = \frac{1}{2}$.

The part of speech tag definitions are: *bos* (*begin sentence marker*), *N* (*noun*), *V* (*verb*), *D* (*determiner*), *P* (*preposition*), *eos* (*end of sentence marker*).

$P(t_i t_{i-2}, t_{i-1})$	t_{i-2}	t_{i-1}	t_i
1	bos	bos	N
$\frac{1}{2}$	bos	N	N
$\frac{1}{2}$	bos	N	V
$\frac{1}{2}$	N	N	V
$\frac{1}{2}$	N	N	P
$\frac{1}{2}$	N	V	D
$\frac{1}{3}$	N	V	V
$\frac{1}{3}$	N	V	P
1	V	D	N
1	V	V	D
1	N	P	D
1	V	P	D
1	P	D	N
1	D	N	eos

$P(w_i t_i)$	t_i	w_i
1	D	an
$\frac{2}{5}$	N	time
$\frac{3}{5}$	N	arrow
$\frac{1}{5}$	N	flies
1	P	like
$\frac{1}{2}$	V	like
$\frac{1}{2}$	V	flies
1	eos	eos
1	bos	bos

- a. Consider a Jelinek-Mercer style interpolation smoothing scheme for $P(w_i | t_i)$:

$$P_{jm}(w_i | t_i) = \Lambda[t_i] \cdot P(w_i | t_i) + (1 - \Lambda[t_i]) \cdot P(w_i)$$

Λ is an array with a value $\Lambda[t_i]$ for each part of speech tag t_i , such that $0 \leq \Lambda[t_i] \leq 1$. Provide a condition on Λ that must be satisfied to ensure that P_{jm} is a well-defined probability model.

Answer: Because of the following fact about $P(w_i | t_i)$:

$$\sum_{w_i} P(w_i | t_i) = 1$$

and in P_{jm} we are given t_i , so to interpolate with $P(w_i)$ for each t_i independently we have the condition that $0 \leq \Lambda[t_i] \leq 1$. Note that we do **not** require the following condition:

$$\sum_{t_i} \Lambda[t_i] = 1$$

- max mark: 2
- contains correct answer: 1 mark
- answer has no extra conditions that are incorrect: 1 mark

- b. Provide a Hidden Markov Model (*hmm*) that uses the trigram part of speech probability $P(t_i | t_{i-2}, t_{i-1})$ as the transition probability $P_{hmm}(s_j | s_k)$ and the probability $P(w_i | t_i)$ as the emission probability $P_{hmm}(w_j | s_j)$.

Important: Provide the *hmm* in the form of two tables as shown below. The first table contains transitions between states in the *hmm* and the transition probabilities and the second table contains the words emitted at each state and the emission probabilities. Do not provide entries with zero probability.

from-state s_k	to-state s_j	$P(s_j s_k)$

state s_j	emission w	$P(w s_j)$

Hint: In your *hmm* the state $\langle N, \text{eos} \rangle$ will have emission of word eos with probability 1 and will not

have transitions to any other states.

Answer: Here are the two tables that define the HMM, the transition table on the left and the emission table on the right:

from-state s_k	to-state s_j	$P(s_j s_k)$
<i>bos, bos</i>	<i>bos, N</i>	$P(N bos, bos)$ 1
<i>bos, N</i>	<i>N, N</i>	$P(N bos, N)$ $\frac{1}{2}$
<i>bos, N</i>	<i>N, V</i>	$P(V bos, N)$ $\frac{1}{2}$
<i>N, N</i>	<i>N, V</i>	$P(V N, N)$ $\frac{1}{2}$
<i>N, N</i>	<i>N, P</i>	$P(P N, N)$ $\frac{1}{2}$
<i>N, V</i>	<i>V, D</i>	$P(D N, V)$ $\frac{1}{3}$
<i>N, V</i>	<i>V, V</i>	$P(V N, V)$ $\frac{1}{3}$
<i>N, V</i>	<i>V, P</i>	$P(P N, V)$ $\frac{1}{3}$
<i>V, D</i>	<i>D, N</i>	$P(N V, D)$ 1
<i>V, V</i>	<i>V, D</i>	$P(D V, V)$ 1
<i>N, P</i>	<i>P, D</i>	$P(D N, P)$ 1
<i>V, P</i>	<i>P, D</i>	$P(D V, P)$ 1
<i>P, D</i>	<i>D, N</i>	$P(N P, D)$ 1
<i>D, N</i>	<i>N, eos</i>	$P(eos D, N)$ 1

state s_j	emission w	$P(w s_j)$
<i>bos, bos</i>	<i>bos</i>	1
<i>bos, N</i>	<i>time</i>	$\frac{1}{2}$
<i>bos, N</i>	<i>arrow</i>	$\frac{1}{2}$
<i>bos, N</i>	<i>flies</i>	$\frac{1}{2}$
<i>N, N</i>	<i>time</i>	$\frac{1}{2}$
<i>N, N</i>	<i>arrow</i>	$\frac{1}{2}$
<i>N, N</i>	<i>flies</i>	$\frac{1}{2}$
<i>N, V</i>	<i>like</i>	$\frac{1}{2}$
<i>N, V</i>	<i>flies</i>	$\frac{1}{2}$
<i>V, D</i>	<i>an</i>	1
<i>V, V</i>	<i>like</i>	$\frac{1}{2}$
<i>V, V</i>	<i>flies</i>	$\frac{1}{2}$
<i>N, P</i>	<i>like</i>	1
<i>V, P</i>	<i>like</i>	1
<i>P, D</i>	<i>an</i>	1
<i>D, N</i>	<i>time</i>	$\frac{1}{2}$
<i>D, N</i>	<i>arrow</i>	$\frac{1}{2}$
<i>D, N</i>	<i>flies</i>	$\frac{1}{2}$

- max mark: 8
- There are 32 entries in the tables above for transition and emission. $\frac{1}{4}$ mark for each table entry.

c. Based on your *hmm* constructed in 2b. what is the state sequence that would be provided by the Viterbi algorithm for the following input sentence:

bos bos time flies like an arrow eos

Answer:

Note that the only ambiguous words are *flies* (could be *N* or *V*) and *like* (could be *V* or *P*) and so all you need to do is compare the scores for the following sub-sequence. The bold-faced outcome wins for this sub-sequence which determines the best state sequence for the entire input.

flies	like	
(N, V)	(V, P)	$\frac{1}{2} \times \frac{1}{3} \times \mathbf{1}$
(N, V)	(V, V)	$\frac{1}{2} \times \frac{1}{3} \times \frac{1}{2}$
(N, N)	(N, P)	$\frac{1}{5} \times \frac{1}{2} \times 1$
(N, N)	(N, V)	$\frac{1}{5} \times \frac{1}{2} \times \frac{1}{2}$

Since the best state sequence is then (bos,N)–(N,V)–(V,P)–(P,D)–(D,N)–(N,eos) the output best state sequence will be *bos/bos, time/N, flies/V, like/P, an/D, arrow/N, eos/eos*.

- max mark: 5
- correct state sequence: 2 marks
- correct table (either small table above or full Viterbi table below): 3 marks

Answer: The full table is given below but you do not need to compute the entire table to solve this question.

bos	time	flies	like	an	arrow	eos	
(bos,bos)	(bos,N)	(N,V)	(V,P)	(P,D)	(D,N)	(N,eos)	
1	$\times 1 \times \frac{2}{5}$	$\times \frac{1}{2} \times \frac{1}{2}$	$\times \frac{1}{3} \times 1$	$\times 1 \times 1$	$\times 1 \times \frac{2}{5}$	$\times 1 \times 1$	$= \frac{1}{75}$
(bos,bos)	(bos,N)	(N,V)	(V,V)	(V,D)	(D,N)	(N,eos)	
1	$\times 1 \times \frac{2}{5}$	$\times \frac{1}{2} \times \frac{1}{2}$	$\times \frac{1}{3} \times \frac{1}{2}$	$\times 1 \times 1$	$\times 1 \times \frac{2}{5}$	$\times 1 \times 1$	$= \frac{1}{150}$
(bos,bos)	(bos,N)	(N,N)	(N,P)	(P,D)	(D,N)	(N,eos)	
1	$\times 1 \times \frac{2}{5}$	$\times \frac{1}{2} \times \frac{1}{5}$	$\times \frac{1}{2} \times 1$	$\times 1 \times 1$	$\times 1 \times \frac{2}{5}$	$\times 1 \times 1$	$= \frac{1}{125}$
(bos,bos)	(bos,N)	(N,N)	(N,V)	(V,D)	(D,N)	(N,eos)	
1	$\times 1 \times \frac{2}{5}$	$\times \frac{1}{2} \times \frac{1}{5}$	$\times \frac{1}{2} \times \frac{1}{2}$	$\times \frac{1}{3} \times 1$	$\times 1 \times \frac{2}{5}$	$\times 1 \times 1$	$= \frac{1}{750}$