midterm-questionbank

04 December 2020 17:07

midterm-questionbank

CS 6320.002: Natural Language Processing
Fall 2020
Midterm
19 Oct. 2020

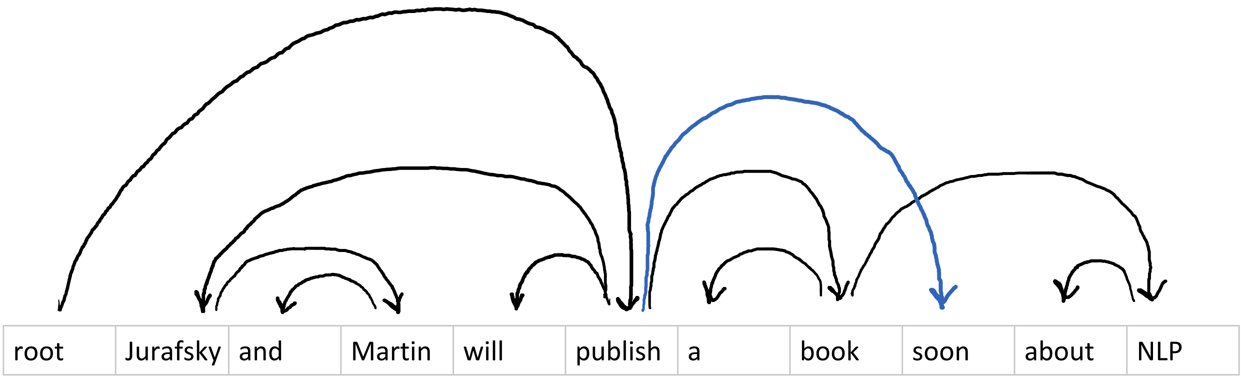
1

"Projectivize" the dependency parse tree of the sentence "Jurafsky and Martin will publish a book soon about NLP" using Nivre and Nilsson's method. What arc(s) would you need to lift, and what new arc(s) would you replace them with? You can give your answer using arrows ($A \rightarrow B$) or ordered pairs (A, B).

2

Suppose we have a dataset of restaurant reviews, and we want to perform binary sentiment classification. However, the restaurant corpus does not come with gold standard labels. We have another dataset consisting of movie reivews that does come with gold standard labels. How could you use the movie corpus, along with clustering, to develop a system to perform sentiment classification on the restaurant corpus? Describe how you would design and train the system (3-4 senctences).

1. Projectivize Jurafsky and Martin will publish a book



2. Given data without label : Restaurant reviews

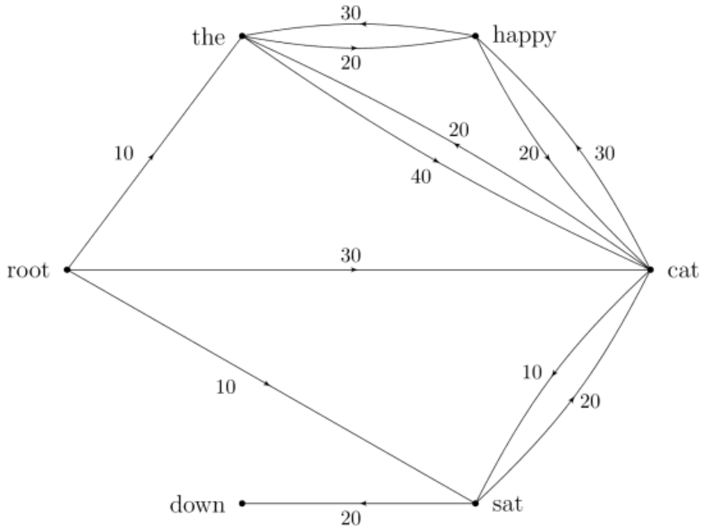
   Task : Sentiment classification

   Data with label : Movie corpus

   Movie corpus along with clustering for restaurant corpus

3

Suppose we have the following graph model of candidate dependencies for the sentence "the happy cat sat down":



You can assume that any edges not shown have score 0. What is the highest-scoring dependency parse tree based on this graph (you can give your answer using arrows (A $\rightarrow$ B) or ordered pairs (A, B)), and what is its total score? Show your work.

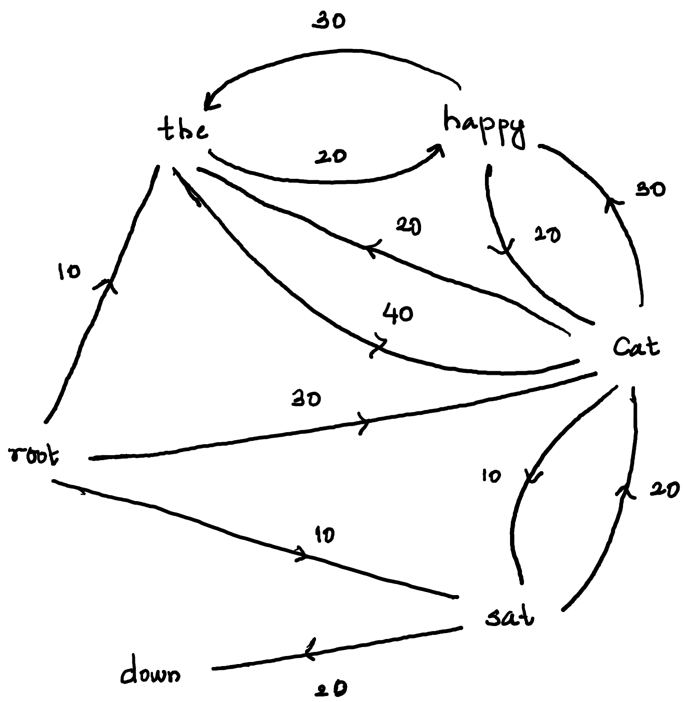
4

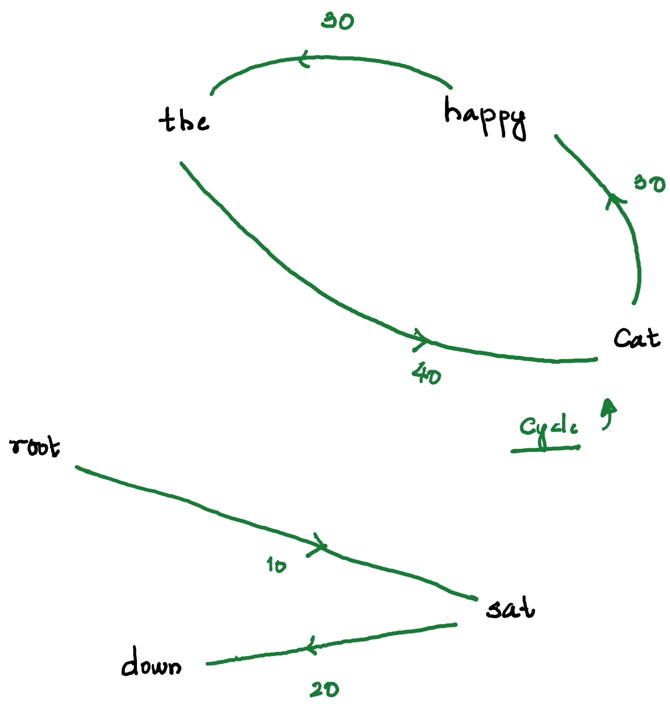
Suppose we train a hidden Markov model to do part-of-speech tagging using the following training sequences:

- the/DT cat/NN is/VB in/PP the/DT box/NN
- the/DT cat/NN sat/VB on/PP the/DT mat/NN
- the/DT cat/NN is/VB black/JJ

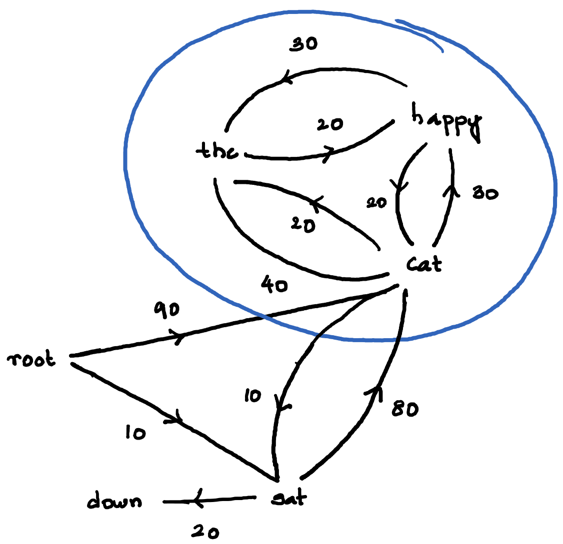
We use a start tag $\langle S \rangle$, but no end tag, bigram tag transitions, and we smooth the transition probabilities using interpolation with $\lambda_1 = \lambda_2 = \frac{1}{2}$. What is the predicted tag sequence for "the black box cat" under this model, and what is its probability? Show your work. You can leave your answer as a product of fractions.

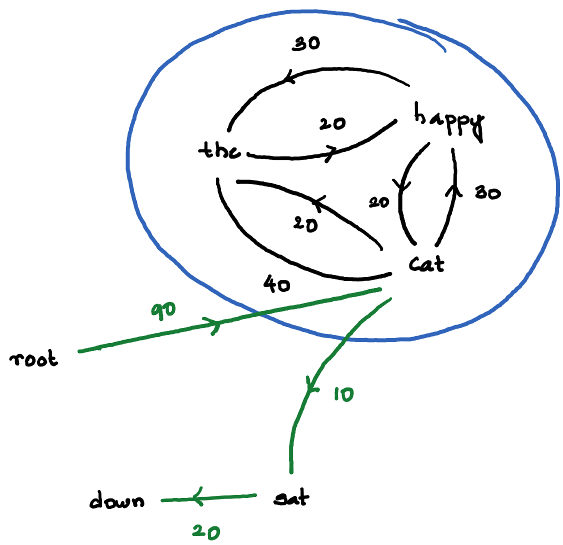
3)

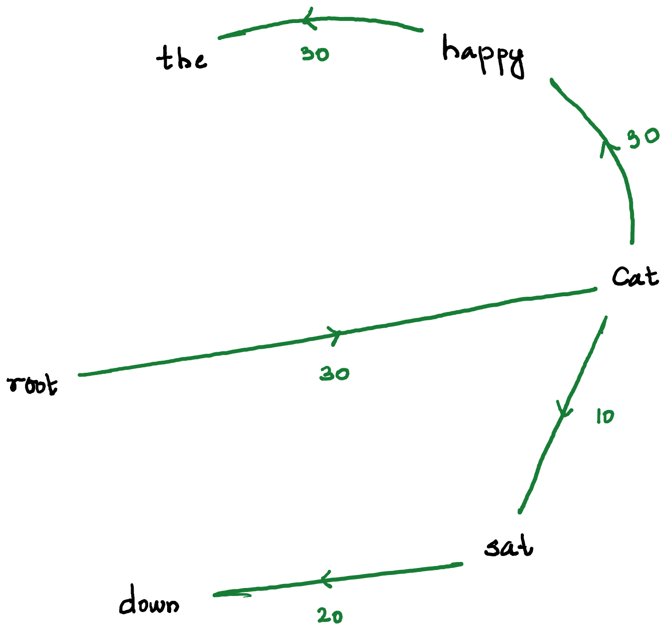




CONDENSED GRAPH :







Total Score : 30 + 30 + 30 + 10 + 20

= 120

4.

<S>	DT	NN	VB	PP	DT	NN	
<S>	DT	NN	VB	PP	DT	NN	
<S>	DT	NN	VB	JJ			

TRANSITION PROBABILITIES

	DT	NN	VB	PP	JJ
<S>	1	0	0	0	0
DT	0	1	0	0	0
NN	0	0	1	0	0
VB	0	0	0	0.67	0.33
PP	1	0	0	0	0
JJ	0	0	0	0	0

UNIGRAM TRANSITION PROBABILITIES

	c	p
<S>	3	3/19 = 0.1579
DT	5	5/19 = 0.2632
NN	5	5/19 = 0.2632
VB	3	3/19 = 0.1579
PP	2	2/19 = 0.1053
JJ	1	1/19 = 0.0526
	19	

SMOOTHED TRANSITION PROBABILITIES

	DT	NN	VB	PP	JJ
<S>	0.578947	0.078947	0.078947	0.078947	0.078947
DT	0.131579	0.631579	0.131579	0.131579	0.131579
NN	0.131579	0.131579	0.631579	0.131579	0.131579
VB	0.078947	0.078947	0.078947	0.413947	0.243947
PP	0.552632	0.052632	0.052632	0.052632	0.052632
JJ	0.026316	0.026316	0.026316	0.026316	0.026316

EMISSION PROBABILITIES

	the	cat	is	in	box	sat	on	mat	black
DT	1	0	0	0	0	0	0	0	0
NN	0	0.6	0	0	0.2	0	0	0.2	0
VB	0	0	0.67	0	0	0.33	0	0	0
PP	0	0	0	0.5	0	0	0.5	0	0
JJ	0	0	0	0	0	0	0	0	1

	the	black	box	cat
DT	0.578			
NN	0			
VB	0			
PP	0			
JJ	0			

	the	black	box	cat
DT	0.578	0		
NN	0	0		
VB	0	0		
PP	0	0		
JJ	0	0.578*0.1315 = 0.076		

	the	black	box	cat
DT	0.578	0	0	
NN	0	0	0.076*0.0263*0.2 = 0.0004	
VB	0	0	0	
PP	0	0	0	
JJ	0	0.076	0	

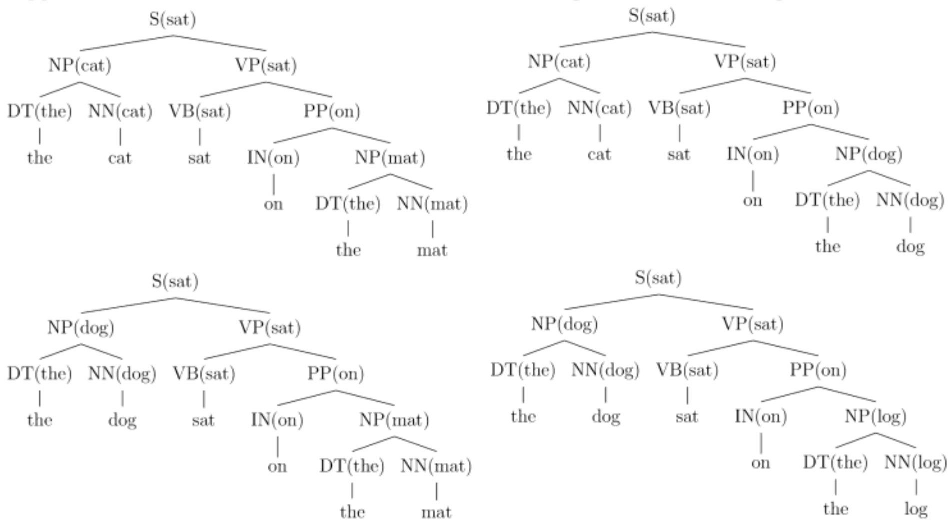
	the	black	box	cat
DT	0.578	0	0	0
NN	0	0	0.0004	0.0004*0.1315*0.6
VB	0	0	0	0
PP	0	0	0	0
JJ	0	0.076	0	0
	DT	JJ	NN	NN

Page 3

04 December 2020 17:07

5

Suppose we want to estimate a lexicalized PCFG using these four training trees:



What is the probability of the rule PP(on) $\rightarrow$ IN(on) NP(log), estimated using Charniak's method with uniform λs for smoothing? You can assume that the position of the head is fixed for each rule, so there is only one rule, not two different rules depending on head position. Show your work.

6

Suppose we have two unigram probability distributions:

$$p(\text{cat}) = \tfrac{1}{2} \quad p(\text{dog}) = \tfrac{1}{4}$$
$$p(\text{mat}) = \tfrac{1}{8} \quad p(\text{log}) = \tfrac{1}{8}$$

and

$$q(\text{cat}) = \tfrac{1}{8} \quad q(\text{dog}) = \tfrac{1}{8}$$
$$q(\text{mat}) = \tfrac{1}{4} \quad q(\text{log}) = \tfrac{1}{2}$$

What is the cross-entropy $H(p, q)$? Show your work.

5. $p(PP(on) \rightarrow IN(on)\ NP(log))$

$$= \underbrace{p(PP(on) \rightarrow IN(on)\ NP \mid PP(on))}_{\text{I}}\ \underbrace{p(log \mid PP(on) \rightarrow IN(on)\ NP)}_{\text{II}}$$

Considering I : $p(PP(on) \rightarrow IN(on)\ NP \mid PP(on))$

$$= \lambda_1 p_{MLE}(PP(on) \rightarrow IN(on)\ NP \mid PP(on)) + \lambda_2 p_{MLE}(PP \rightarrow IN\ NP \mid PP)$$

We have $\lambda_1 = \lambda_2 = 1/2$

$$p_{MLE}(PP(on) \rightarrow IN(on)\ NP \mid PP(on)) = 1$$

$$p_{MLE}(PP \rightarrow IN\ NP \mid PP) = 1$$

For I we have $p(PP(on) \rightarrow IN(on)\ NP \mid PP(on)) = 1$

Considering II : $p(log \mid PP(on) \rightarrow IN(on)\ NP)$

$$= \lambda_1 p_{MLE}(log \mid PP(on) \rightarrow IN(on)\ NP) + \lambda_2 p_{MLE}(log \mid PP \rightarrow IN\ NP) + \lambda_3 p_{MLE}(log \mid NP)$$

We have $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$, $p_{MLE}(log \mid PP(on) \rightarrow IN(on)\ NP) = \frac{1}{4}$

$p_{MLE}(log \mid PP \rightarrow IN\ NP) = \frac{1}{4}$, $p_{MLE}(log \mid NP) = \frac{1}{8}$

For II : $\frac{1}{3}\left[\frac{1}{4} + \frac{1}{4} + \frac{1}{8}\right] = \frac{1}{3}\left[\frac{4}{8} + \frac{1}{8}\right] = \frac{5}{24}$

$\therefore\ p(PP(on) \rightarrow IN(on)\ NP(dog)) = \boxed{\frac{5}{24}}$

6)

$$H(p,q) = E_p\left[-\log(q)\right]$$

$$= -\left[\frac{1}{2}\log\left(\frac{1}{8}\right) + \frac{1}{4}\log\left(\frac{1}{8}\right) + \frac{1}{8}\log\left(\frac{1}{4}\right) + \frac{1}{8}\log\left(\frac{1}{2}\right)\right]$$

$$= -\left[\frac{1}{2}(-3) + \frac{1}{4}(-3) + \frac{1}{8}(-2) + \frac{1}{8}(-1)\right]$$

$$= \underbrace{\frac{3}{2} + \frac{3}{4}} + \frac{1}{4} + \frac{1}{8}$$

$$= \frac{9}{4} + \frac{1}{4} + \frac{1}{8} = \frac{20}{8} + \frac{1}{8} = \frac{21}{8}$$

7

Suppose we have a training corpus of two sentences:

1.	grand	cheval	2.	grand
	big	horse		big

Now suppose we have a word-level statistical machine translation model with the parameters $q(j|i, l, m)$ and $t(f_i|e_j)$ initialized uniformly. Perform one iteration of expectation maximization to update the parameters using the training corpus. What is $t(\text{grand}|\text{horse})$ at the end of this iteration? Show your work.

8

Suppose we have a training corpus of two sentences:

1.	souris	blanche	2.	souris
	white	mouse		mouse

Additionally, suppose we already have a trained IBM Model 1 statistical machine translation system, and now we want to train an IBM Model 2. The parameters $q(j|i, l, m)$ are initialized uniformly, and the parameters $t(f_i|e_j)$ are initialized using the Model 1 as follows:

- $t(\text{souris}|\text{white}) = \frac{1}{2}$
- $t(\text{souris}|\text{mouse}) = \frac{3}{4}$
- $t(\text{blanche}|\text{white}) = \frac{1}{2}$
- $t(\text{blanche}|\text{mouse}) = \frac{1}{4}$

Perform one iteration of expectation maximization to update the parameters using the training corpus. What is $t(\text{blanche}|\text{mouse})$ at the end of this iteration? Show your work.

9

Suppose we have a language model that uses Kneser-Ney smoothing to handle rare/out-of-vocabulary words. Is this strategy for handling rare/out-of-vocabulary words sufficient, or do we also need to use an unknown word token? Briefly explain your answer (1-2 sentences).

7. EM to update parameters and find t(grand|horse)

<u>INITIAL PARAMETERS: [Initialized uniformly]</u>

t	grand	cheval
big	1/2	1/2
horse	1/2	1/2

We have two sentences :

For k = 1 : (m = 2, l=2)
For k = 2 : (m = 1, l=1)

q(j|i,l,m)
j - ENGLISH WORD
i - FRENCH WORD

q(1|1,1,1)	1		
q(1|1,2,2)	1/2	q(2|1,2,2)	1/2
q(1|2,2,2)	1/2	q(2|2,2,2)	1/2

<u>NEXT STEP:</u>

δ(k,i,j)

δ(2,1,1)	1	
δ(1,1,1)	q(1|1,2,2) t(grand|big) / [(q(1|1,2,2) t(grand|big) + q(2|1,2,2) t(grand|horse))] = (1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,1,2)	q(2|1,2,2) t(grand|horse) / [(q(1|1,2,2) t(grand|big) + q(2|1,2,2) t(grand|horse))] =(1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,2,1)	q(1|2,2,2) t(cheval|big) / [q(1|2,2,2) t(cheval|big) + q(2|2,2,2) t(cheval|horse))] = (1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,2,2)	q(2|2,2,2) t(cheval|horse) / [q(1|2,2,2) t(cheval|big) + q(2|2,2,2) t(cheval|horse))] =(1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2

c(big, grand)	1/2 + 1		c(big)	2		t(grand|horse)	c(horse,grand)/c(horse)
c(big, cheval)	1/2		c(horse)	1			1/2
c(horse, grand)	1/2						
c(horse, cheval)	1/2						

8. EM to update parameters and find t(blanche|mouse)

INITIAL PARAMETERS:

t	souris	blanche
white	1/2	1/2
mouse	3/4	1/4

We have two sentences :

For k = 1 : (m = 2, l=2)
For k = 2 : (m = 1, l=1)

Initial q's are initialized uniformly

q(j|i,l,m)
j - ENGLISH WORD
i - FRENCH WORD

q(1|1,1,1)	1		
q(1|1,2,2)	1/2	q(2|1,2,2)	1/2
q(1|2,2,2)	1/2	q(2|2,2,2)	1/2

NEXT STEP:

δ(k,i,j)
j - ENGLISH WORD
i - FRENCH WORD

δ(2,1,1)	1	
δ(1,1,1)	q(1|1,2,2) t(souris|white) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] = (1/2)(1/2)/ [(1/2)(1/2) + (1/2)(3/4)]	2/5
δ(1,1,2)	q(2|1,2,2) t(souris|mouse) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] =(1/2)(3/4)/ [(1/2)(1/2) + (1/2)(3/4)]	3/5
δ(1,2,1)	q(1|2,2,2) t(blanche|white) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/2) / [(1/2)(1/2) + (1/2)(1/4)]	2/3
δ(1,2,2)	q(2|2,2,2) t(blanche|mouse) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/4) / [(1/2)(1/2) + (1/2)(1/4)]	1/3

c(white, souris)	2/5		c(white)	2/5 + 2/3		t(blanche|mouse)	c(mouse,blanche)/c(mouse)
c(white, blanche)	2/3		c(mouse)	3/5 + 1/3 + 1			(1/3) / [3/5 + 1/3 + 1]
c(mouse, souris)	3/5 + 1						(1/3) / [3/5 + 4/3]
c(mouse, blanche)	1/3						5/29

10

Suppose we have a trained first-order hidden Markov model for part-of-speech tagging with the parameters below (assume the value is 0 if the parameter is not shown in the tables).

Tag transition probabilities:

	DT	JJ	NN	VB
⟨S⟩	0.5	0.1	0.3	0.1
DT	0.0	0.5	0.5	0.0
JJ	0.0	0.4	0.6	0.0
NN	0.1	0.0	0.2	0.7
VB	0.4	0.2	0.2	0.2

Word emission probabilities:

	the	old	man	ships
DT	1.0	0.0	0.0	0.0
JJ	0.0	0.8	0.2	0.0
NN	0.0	0.2	0.4	0.4
VB	0.0	0.0	0.5	0.5

What is the predicted tag sequence for the sentence "the old man the ships," decoded using beam search Viterbi with beam size 2, and what is the predicted probability of that tag sequence? Show your work. You can leave the answer as a product of decimals.

11

Which of the following underlined phrases is a constituent? Choose all that apply.

A These black cats detest those <u>green peas</u>.
B <u>These black cats</u> detest those green peas.
C Put it over <u>on the table</u> that's by the door.
D Put it <u>over on the table</u> that's by the door.

12

Suppose we have a trained first-order hidden Markov model for part-of-speech tagging with the parameters below (assume the value is 0 if the parameter is not shown in the tables).

Tag transition probabilities:

	DT	NN	VB	⟨/S⟩
⟨S⟩	0.5	0.3	0.2	0.0
DT	0.0	0.9	0.1	0.0
NN	0.1	0.2	0.3	0.4
VB	0.4	0.2	0.2	0.2

Word emission probabilities:

	the	can	see	⟨/s⟩
DT	1.0	0.0	0.0	0.0
NN	0.0	0.9	0.1	0.0
VB	0.0	0.5	0.5	0.0
⟨/S⟩	0.0	0.0	0.0	1.0

What is the predicted tag sequence for the sentence "can the can see ⟨/s⟩", decoded using beam search Viterbi with beam size 2, and what is the predicted probability of that tag sequence? Show your work. You can leave the answer as a product of decimals.

10. The old man the ships - viterbi beam 2

	DT	JJ	NN	VB	CHECK			the	old	man	ships
<S>	0.5	0.1	0.3	0.1	1		DT	1	0	0	0
DT	0	0.5	0.5	0	1		JJ	0	0.8	0.2	0
JJ	0	0.4	0.6	0	1		NN	0	0.2	0.4	0.4
NN	0.1	0	0.2	0.7	1		VB	0	0	0.5	0.5
VB	0.4	0.2	0.2	0.2	1						

STEP - 1

	the	old	man	the	ships
DT	(0.5) (1)				
JJ	(0.1) (0)				
NN	(0.1) (0)				
VB	(0.1) (0)				

	the	old	man	the	ships
DT	0.5	0			
JJ	0	0.5*0.5*0.8			
NN	0	0.5*0.5*0.2			
VB	0	0			

	the	old	man	the	ships
DT	0.5	0	0		
JJ	0	0.2	Max(0.4*0.2*0.2, 0) = Max(0.016, 0)		
NN	0	0.05	Max(0.6*0.4*0.2, 0.2*0.4*0.05) = Max(0.048, 0.004)		
VB	0	0	Max(0, 0.7*0.5*0.05) = Max(0, 0.0175)		

	the	old	man	the	ships
DT	0.5	0	0	Max(0.1*0.048, 0.4*0.0175) = Max(0, 0.0048, 0.07)	
JJ	0	0.2	0	0	
NN	0	0.05	0.048	0	
VB	0	0	0.0175	0	

	the	old	man	the	ships
DT	0.5	0	0	0.07	0
JJ	0	0.2	0	0	0
NN	0	0.05	0.048	0	0.5*0.4*0.07 = 0.014
VB	0	0	0.0175	0	0

DT	NN	VB	DT	NN

11. Constituents

A These black cats detest those <u>green peas.</u>
B <u>These black cats</u> detest those green peas.
C Put it over <u>on the table</u> that's by the door.
D Put it <u>over on the table</u> that's by the door.

A. Not a constituent
B. Constituent
C. Not a Constituent
D. Not a Constituent

12. Can the can see </S> - viterbi beam 2

	DT	NN	VB	</S>
<S>	0.5	0.3	0.2	0
DT	0	0.9	0.1	0
NN	0.1	0.2	0.3	0.4
VB	0.4	0.2	0.2	0.2

	the	can	see	</S>
DT	1	0	0	0
NN	0	0.9	0.1	0
VB	0	0.5	0.5	0
</S>	0	0	0	1

	can	the	can	see	</s>
DT	0				
NN	0.3*0.9				
VB	0.2*0.5				
</S>	0				

	can	the	can	see	</s>
DT	0	Max(0.1*1*0.27, 0.4*1*0.1) = Max(0.027, 0.04)			
NN	0.27	0			
VB	0.1	0			
</S>	0	0			

	can	the	can	see	</s>
DT	0	0.04	0		
NN	0.27	0	0.9*0.9*0.04		
VB	0.1	0	0.1*0.5*0.04		
</S>	0	0	0		

	can	the	can	see	</s>
DT	0	0.04	0	0	
NN	0.27	0	0.0324	Max(0.2*0.1*0.0324, 0.2*0.1*0.002) = Max(0.000648, 0.00004)	
VB	0.1	0	0.002	Max(0.3*0.5*0.0324, 0.2*0.5*0.002) = Max(0.00486, 0.0002)	
</S>	0	0	0	0	

	can	the	can	see	</s>
DT	0	0.04	0	0	0
NN	0.27	0	0.0324	0.000648	0
VB	0.1	0	0.002	0.00486	0
</S>	0	0	0	0	Max(0.4*0.000648, 0.2*0.00486)= Max(0.0002592, 0.000972)

	can	the	can	see	</s>
DT	0	0.04	0	0	0
NN	0.27	0	0.0324	0.000648	0
VB	0.1	0	0.002	0.00486	0
</S>	0	0	0	0	0.000972

VB	DT	NN	VB	</S>

13

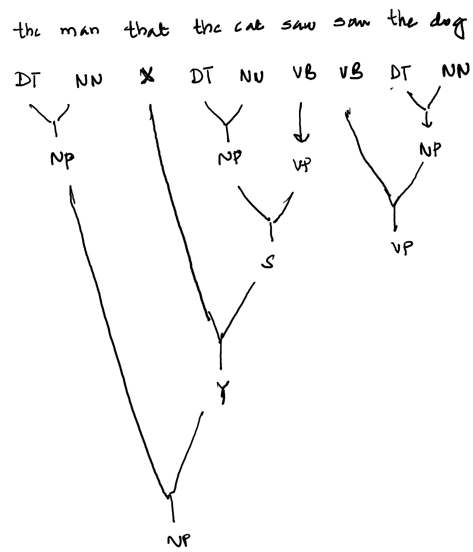
Suppose you want to build a system to perform phrase-based machine translation, and you have access to a working word sense disambiguation (WSD) system. How could you use the information provided by the WSD system to try to improve the performance of your model? Which parameter(s) of the model would you modify, and how? (2-3 sentences.)

14

Suppose we have the following context-free grammar:

S	$\rightarrow$	NP VP
VP	$\rightarrow$	VB
VP	$\rightarrow$	VB NP
NP	$\rightarrow$	DT NN
DT	$\rightarrow$	the
NN	$\rightarrow$	cat
NN	$\rightarrow$	dog
NN	$\rightarrow$	man
VB	$\rightarrow$	saw

What new rule(s) would you have to add to the grammar to be able to parse the sentence "the man that the cat saw saw the dog"? If you need new symbols, use X, Y, and Z.



15

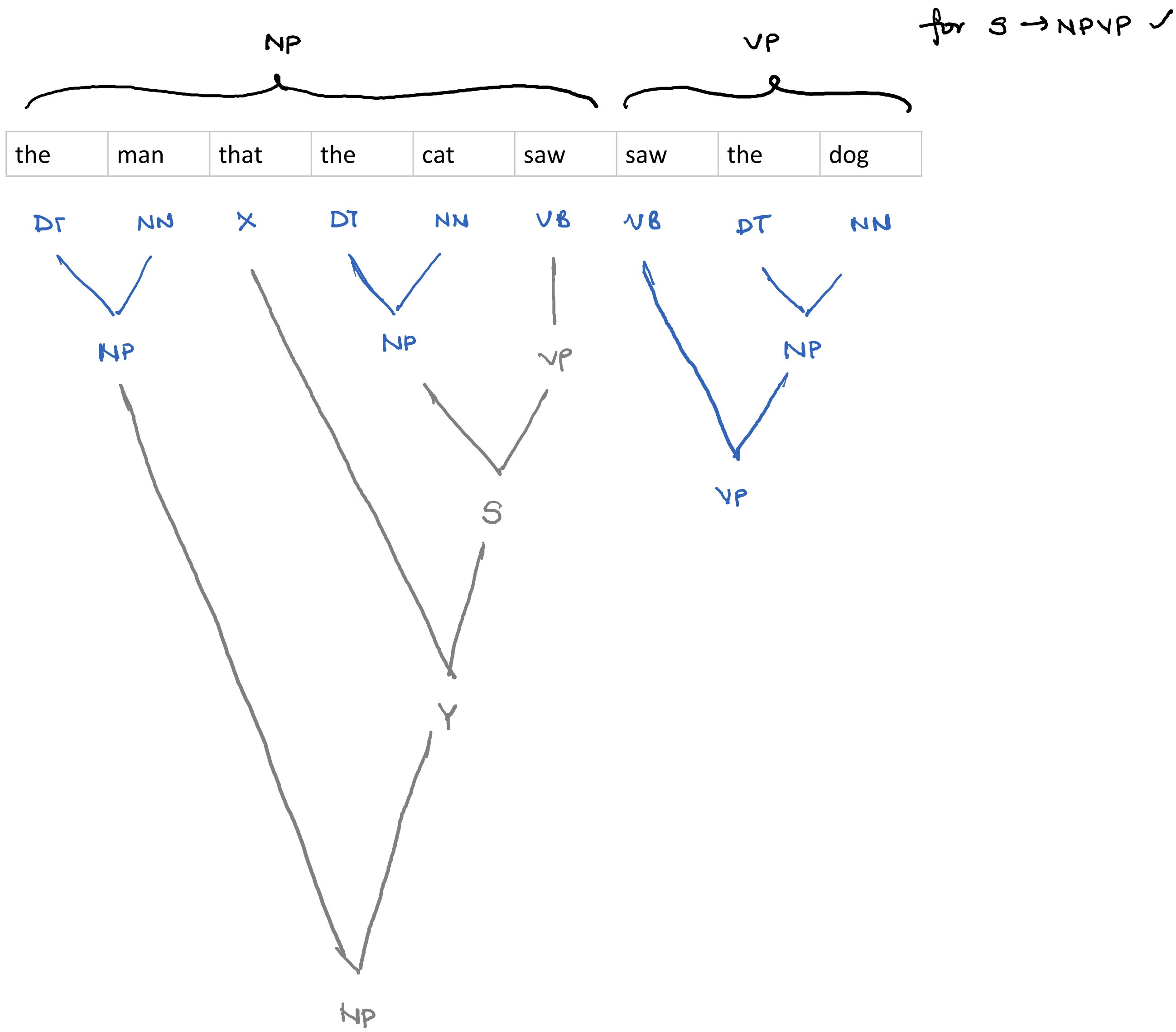
Suppose we train a Naive Bayes classifier on the following training sequences (the letter after the comma is the class label):

- the cat sat on the mat, A
- the cat sat in the hat, A
- the dog sat on the log, B
- the dog sat on the cat, B
- the fish sat in the dish, C
- the fish in the hat sat, C

Our classifier uses skipgram count features, no start or end tokens, and Laplace smoothing with $\delta = 1$. Given the test sentence "the cat in the hat," what is the predicted probability of this sentence belonging to class C under this model? Show your work. You can leave the answer as a product of fractions.

14. New CFG rules to be added

X --> that
Y --> X S
NP --> NP Y



15. Naive Bayes

1. the cat sat on the mat, A
2. the cat sat in the hat, A
3. the dog sat on the log, B
4. the dog sat on the cat, B
5. the fish sat in the dish, C
6. the fish in the hat sat, C

Skipgram features
Laplace smoothing

Skipgram : We also include counts of skipgrams

A :
(the, sat) - 2, (cat, on) - 1, (sat, the) - 2, (on, mat) - 1, (cat, in) - 1, (in, hat) - 1

B :
(the,sat) - 2, (dog, on) - 2, (sat, the) - 2, (on, log) - 1, (on, cat) - 1

C:
(the, sat) - 2, (the, in) - 1, (fish, in) - 1, (fish, the) - 1, (sat, the) - 1, (in, hat) - 1, (in, dish) - 1

A : COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	2	2	0	1	0	1	0	0	0	0	6
cat	0	0	2	1	0	1	0	0	0	0	0	4
sat	2	0	0	1	0	1	0	0	0	0	0	4
on	1	0	0	0	1	0	0	0	0	0	0	2
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	1	0	0	0	0	0	1	0	0	0	0	2
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	2	4	2	2	2	2	0	0	0	0	18

A : SMOOTHED COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	3	3	1	2	1	2	1	1	1	1	17
cat	1	1	3	2	1	2	1	1	1	1	1	15
sat	3	1	1	2	1	2	1	1	1	1	1	15
on	2	1	1	1	2	1	1	1	1	1	1	13
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	2	1	1	1	1	1	2	1	1	1	1	13
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	13	15	13	13	13	13	11	11	11	11	139

A : PROBABILITY MATRIX

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.00763	0.0229	0.0229	0.0076	0.0153	0.0076	0.0153	0.00763359	0.0076	0.007633588	0.0076	0.1298
cat	0.00763	0.0076	0.0229	0.0153	0.0076	0.0153	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.1145
sat	0.0229	0.0076	0.0076	0.0153	0.0076	0.0153	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.1145
on	0.01527	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.0992
mat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
in	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.00763359	0.0076	0.007633588	0.0076	0.0992
hat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
dog	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
log	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
fish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
dish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
	0.1145	0.0992	0.1145	0.0992	0.0992	0.0992	0.0992	0.08396947	0.084	0.083969466	0.084	1.0611

the	cat	in	the	hat						
	0.0229	0.0153	0.0153	0.0153	8.1494E-08	0.3333	2.71647E-08		SUM	3.62196E-08
						A	0.75			

B: COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	1	2	0	0	0	0	2	1	0	0	6
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	2	0	0	2	0	0	0	0	0	0	0	4
on	2	1	0	0	0	0	0	0	1	0	0	4
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	0	0	0	0	0	0	0	0	0	0	0	0
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	2	2	0	0	0	0	0	0	0	4
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	2	4	4	0	0	0	2	2	0	0	18

B: SMOOTHED COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	2	3	1	1	1	1	3	2	1	1	17
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	3	1	1	3	1	1	1	1	1	1	1	15
on	3	2	1	1	1	1	1	1	2	1	1	15
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	1	1	1	1	1	1	1	1	1	1	1	11
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	3	3	1	1	1	1	1	1	1	15
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	13	15	15	11	11	11	13	13	11	11	139

B: PROBABILITIES

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0153	0.0229	0.0076	0.0076	0.0076	0.0076	0.0229	0.0153	0.0076	0.0076	0.1298
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0229	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
on	0.0229	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.1145
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
hat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dog	0.0076	0.0076	0.0229	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.1145	0.0992	0.1145	0.1145	0.084	0.084	0.084	0.0992	0.0992	0.084	0.084	1.0611

the	cat	in	the	hat			
	0.0153	0.0076	0.0076	0.0076	7E-09	0.3333	2E-09
						B	0.0625

C : COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	0	2	0	0	1	1	0	0	2	1	7
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	1	0	0	0	0	1	0	0	0	0	0	2
on	0	0	0	0	0	0	0	0	0	0	0	0
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	2	0	0	0	0	0	1	0	0	0	1	4
hat	0	0	1	0	0	0	0	0	0	0	0	1
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	1	0	1	0	0	2	0	0	0	0	0	4
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	4	0	0	4	2	0	0	2	2	18

C : SMOOTHED COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	1	3	1	1	2	2	1	1	3	2	18
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	2	1	1	1	1	2	1	1	1	1	1	13
on	1	1	1	1	1	1	1	1	1	1	1	11
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	3	1	1	1	1	1	2	1	1	1	2	15
hat	1	1	2	1	1	1	1	1	1	1	1	12
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	2	1	2	1	1	3	1	1	1	1	1	15
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	11	15	11	11	15	13	11	11	13	13	139

C : PROBABILITY MATRIX

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0076	0.0229	0.0076	0.0076	0.0153	0.0153	0.0076	0.0076	0.0229	0.0153	0.1374
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0153	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
on	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0153	0.1145
hat	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
dog	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0153	0.0076	0.0153	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.1145	0.084	0.1145	0.084	0.084	0.1145	0.0992	0.084	0.084	0.0992	0.0992	1.0611

the	cat	in	the	hat	Product	P(C)	Product*P(C)
	0.0076	0.0076	0.0229	0.0153	2E-08	0.3333	7E-09
						P(d|c)P(c)/P(d)	**0.1875**

16

Suppose we have the following probabilistic context-free grammar:

S	$\rightarrow$	VB SBAR	1.0
SBAR	$\rightarrow$	NP VP	1.0
VP	$\rightarrow$	VB NP	0.7
NP	$\rightarrow$	DT NN	1.0
DT	$\rightarrow$	the	1.0
NN	$\rightarrow$	can	0.9
NN	$\rightarrow$	see	0.1
VB	$\rightarrow$	can	0.5
VB	$\rightarrow$	see	0.5

How many possible parses are there for the sentence "see the can can the can" under this grammar? Which has the highest probability, and what is that probability? Show your work.

17

Which of the following underlined phrases is a constituent? Choose all that apply.

A They arrived at the concert <u>more quickly than they had expected.</u>
B Ali Baba returned from his travels <u>wiser than before</u>.
C I am <u>very fond</u> of my nephew.

18

Suppose we have the following probabilistic context-free grammar:

NP	$\rightarrow$	DT NBAR	1.0
NBAR	$\rightarrow$	NN	0.4
NBAR	$\rightarrow$	JJ NBAR	0.3
NBAR	$\rightarrow$	NBAR NBAR	0.3
DT	$\rightarrow$	the	1.0
JJ	$\rightarrow$	purple	0.5
JJ	$\rightarrow$	terrible	0.5
NN	$\rightarrow$	people	0.7
NN	$\rightarrow$	eater	0.3

How many possible parses are there for the phrase "the terrible purple people eater" under this grammar? Which has the highest probability, and what is that probability? Show your work.

16.

	see	the	can	can	the	can
see	NN 0.1, VB 0.5	-				
the		DT 1.0	NP 0.9			
can			NN 0.9, VB 0.5	-		
can				NN 0.9, VB 0.5	-	
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	see	the	can	can	the	can
see	NN 0.1, VB 0.5	-	VP 0.7*0.5*0.9=0.315			
the		DT 1.0	NP 0.9	-		
can			NN 0.9, VB 0.5	-	-	
can				NN 0.9, VB 0.5	-	VP 0.7*0.5*0.9=0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	see	the	can	can	the	can
see	NN 0.1, VB 0.5	-	VP 0.315	-		
the		DT 1.0	NP 0.9	-	-	
can			NN 0.9, VB 0.5	-	-	-
can				NN 0.9, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	see	the	can	can	the	can
see	NN 0.1, VB 0.5	-	VP 0.315	-	-	
the		DT 1.0	NP 0.9	-	-	SBAR 0.9*0.315 = 0.2835
can			NN 0.9, VB 0.5	-	-	-
can				NN 0.9, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	see	the	can	can	the	can
see	NN 0.1, VB 0.5	-	VP 0.315	-	-	S 0.5*0.2835 = 0.1418
the		DT 1.0	NP 0.9	-	-	SBAR 0.2835
can			NN 0.9, VB 0.5	-	-	-
can				NN 0.9, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

17. Constituent

A They arrived at the concert more quickly than they had expected.
B Ali Baba returned from his travels wiser than before.
C I am very fond of my nephew.

A. Constituent
B. Constituent
C. Not a constituent

18.

Given CFG is not in CNF.

Converting the rules into CNF by following the procedure gives us:

NP --> DT NBAR 1.0
NBAR --> NN 0.4
NBAR --> JJ NBAR 0.3
NBAR --> NBAR NBAR 0.3
DT --> the 1.0
JJ --> purple 0.5
JJ --> terrible 0.5
NN --> people 0.7
NN --> eater 0.3

1. Removing nullable variables
2. Removing unit variables

NP --> DT NBAR 1.0
NBAR --> NN 0.4
NBAR --> JJ NBAR 0.3
NBAR --> NBAR NBAR 0.3
DT --> the 1.0
JJ --> purple 0.5
JJ --> terrible 0.5
NN --> people 0.7
NN --> eater 0.3

NBAR --> people 0.4*0.7 = 0.28
NBAR --> eater 0.3*0.4 = 0.12

3. Removing offending variables

NP --> DT NBAR 1.0
NBAR --> JJ NBAR 0.3
NBAR --> NBAR NBAR 0.3
DT --> the 1.0
JJ --> purple 0.5
JJ --> terrible 0.5
NN --> people 0.7
NN --> eater 0.3
NBAR --> people 0.28
NBAR --> eater 0.12

	the	terrible	purple	people	eater
the	DT 1.0	-			
terrible		JJ 0.5	-		
purple			JJ 0.5	NBAR 0.3*0.5*0.28 = 0.042	
people				NN 0.7, NBAR 0.28	NBAR 0.3*0.28*0.12=0.0101
eater					NN 0.3, NBAR 0.12

	the	terrible	purple	people	eater
the	DT 1.0	-	-		
terrible		JJ 0.5	-	NBAR 0.3*0.5*0.042 =0.0063	
purple			JJ 0.5	NBAR 0.042	NBAR 0.3*0.5*0.0101=0.0015, NBAR 0.3* 0.042*0.12 = 0.0015
people				NN 0.7, NBAR 0.28	NBAR 0.0101
eater					NN 0.3, NBAR 0.12

	the	terrible	purple	people	eater
the	DT 1.0	-	-	NP 0.0063	
terrible		JJ 0.5	-	NBAR 0.0063	NBAR 0.3*0.5*0.0015 = 0.0002, NBAR 0.3*0.5*0.0015=0.0002, NBAR 0.3*0.0063*0.12 = 0.0002
purple			JJ 0.5	NBAR 0.042	NBAR 0.0015, NBAR 0.0015
people				NN 0.7, NBAR 0.28	NBAR 0.0101
eater					NN 0.3, NBAR 0.12

	the	terrible	purple	people	eater
the	DT 1.0	-	-	NP 0.0063	NP, NP, NP 0.0002
terrible		JJ 0.5	-	NBAR 0.0063	NBAR 0.0002,NBAR 0.0002, NBAR 0.0002
purple			JJ 0.5	NBAR 0.042	NBAR 0.0015, NBAR 0.0015
people				NN 0.7, NBAR 0.28	NBAR 0.0101
eater					NN 0.3, NBAR 0.12

There are 3 possible parses for this sentence

19

Suppose we train a language model on the following training sequences:

- the man saw the dog with the telescope
- the dog saw the man in the park
- the man with the telescope saw the park

We use linear interpolation between a bigram model and a unigram model, with $\lambda_1 = \lambda_2 = 0.5$. The bigram model is smoothed using Kneser-Ney smoothing with $\beta = 1$, and the unigram model is not smoothed. We do not use any start or end tokens. What is $p(\text{with}|\text{man})$ under this model? Show your work. You can leave the answer as a sum of fractions.

20

Suppose you want to build a system to perform word-based machine translation, and you have access to a working word sense disambiguation (WSD) system. How could you use the information provided by the WSD system to try to improve the performance of your model? Which parameter(s) of the model would you modify, and how? (2-3 sentences.)

21

Suppose we have a dataset of restaurant reviews, and we want to perform binary sentiment classification. However, the restaurant corpus does not come with gold standard labels. We have another dataset consisting of movie reivews that does come with gold standard labels. How could you use the movie corpus, along with distant supervision, to develop a system to perform sentiment classification on the restaurant corpus? Describe how you would design and train the system (3-4 senctences).

22

Suppose we train a Naive Bayes classifier on the following training sequences (the letter after the comma is the class label):

- the cat sat on the mat, A
- the cat sat in the hat, A
- the dog sat on the log, B
- the dog sat on the cat, B
- the fish sat in the dish, C
- the fish in the hat sat, C

Our classifier uses skipgram count features, no start or end tokens, and Laplace smoothing with $\delta = 1$. Given the test sentence "the cat in the hat," what is the predicted probability of this sentence belonging to class B under this model? Show your work. You can leave the answer as a product of fractions.

19.

Suppose we train a language model on the following training sequences:
the man saw the dog with the telescope
the dog saw the man in the park
the man with the telescope saw the park

p(with|man)

Linear interpolation :
1. Bigram + Knesser-Ney smoothing (β=1)
2. Unigram Model

c(u,v)	the	man	saw	dog	with	telescope	in	park	SUM	c(u,v)>0
the	0	3	0	2	0	2	0	2	9	4
man	0	0	1	0	1	0	1	0	3	3
saw	3	0	0	0	0	0	0	0	3	1
dog	0	0	1	0	1	0	0	0	2	2
with	2	0	0	0	0	0	0	0	2	1
telescope	0	0	1	0	0	0	0	0	1	1
in	1	0	0	0	0	0	0	0	1	1
park	0	0	0	0	0	0	0	0	0	0
									21	13

<u>DISCOUNTED COUNTS</u>

										$\lambda(w_{i-1})$
c(u,v)	the	man	saw	dog	with	telescope	in	park		
the	0	2	0	1	0	1	0	1		0.44444
man	0	0	0	0	0	0	0	0		**1**
saw	2	0	0	0	0	0	0	0		0.33333
dog	0	0	0	0	0	0	0	0		1
with	1	0	0	0	0	0	0	0		0.5
telescope	0	0	0	0	0	0	0	0		0.5
in	0	0	0	0	0	0	0	0		1
park	0	0	0	0	0	0	0	0		0

P(with|man)

λ(man) = 1

	c(u)	p
the	9	0.38
man	3	0.13
saw	3	0.13
dog	2	0.08
with	2	0.08
telescope	2	0.08
in	1	0.04
park	2	0.08
	24	1

P(with|man)

= 0.5*[Ps(with|man)] + 0.5*P(with)
= 0.5*[c_discounted(man, with)/c(man) + λ(man) * P_cont(dog)] + 0.5*(2/24)

c_discounted(man, with) = 1-1 = 0

λ(man) = 1
P_cont(with) = 2/13

= 0.5* [2/13] + 0.5 * (2/24)

22. Naive Bayes

1. the cat sat on the mat, A
2. the cat sat in the hat, A
3. the dog sat on the log, B
4. the dog sat on the cat, B
5. the fish sat in the dish, C
6. the fish in the hat sat, C

Skipgram features
Laplace smoothing

Skipgram : We also include counts of skipgrams

A :
(the, sat) - 2, (cat, on) - 1, (sat, the) - 2, (on, mat) - 1, (cat, in) - 1, (in, hat) - 1

B :
(the,sat) - 2, (dog, on) - 2, (sat, the) - 2, (on, log) - 1, (on, cat) - 1

C:
(the, sat) - 2, (the, in) - 1, (fish, in) - 1, (fish, the) - 1, (sat, the) - 1, (in, hat) - 1, (in, dish) - 1

A : COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	2	2	0	1	0	1	0	0	0	0	6
cat	0	0	2	1	0	1	0	0	0	0	0	4
sat	2	0	0	1	0	1	0	0	0	0	0	4
on	1	0	0	0	1	0	0	0	0	0	0	2
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	1	0	0	0	0	0	1	0	0	0	0	2
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	2	4	2	2	2	2	0	0	0	0	18

A : SMOOTHED COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	3	3	1	2	1	2	1	1	1	1	17
cat	1	1	3	2	1	2	1	1	1	1	1	15
sat	3	1	1	2	1	2	1	1	1	1	1	15
on	2	1	1	1	2	1	1	1	1	1	1	13
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	2	1	1	1	1	1	2	1	1	1	1	13
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	13	15	13	13	13	13	11	11	11	11	139

A : PROBABILITY MATRIX

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.00763	0.0229	0.0229	0.0076	0.0153	0.0076	0.0153	0.00763359	0.0076	0.007633588	0.0076	0.1298
cat	0.00763	0.0076	0.0229	0.0153	0.0076	0.0153	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.1145
sat	0.0229	0.0076	0.0076	0.0153	0.0076	0.0153	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.1145
on	0.01527	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.0992
mat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
in	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.00763359	0.0076	0.007633588	0.0076	0.0992
hat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
dog	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
log	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
fish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
dish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
	0.1145	0.0992	0.1145	0.0992	0.0992	0.0992	0.0992	0.08396947	0.084	0.083969466	0.084	1.0611

the	cat	in	the	hat						
	0.0229	0.0153	0.0153	0.0153	8.1494E-08	0.3333	2.71647E-08		SUM	3.62196E-08
						A	0.75			

B: COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	1	2	0	0	0	0	2	1	0	0	6
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	2	0	0	2	0	0	0	0	0	0	0	4
on	2	1	0	0	0	0	0	0	1	0	0	4
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	0	0	0	0	0	0	0	0	0	0	0	0
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	2	2	0	0	0	0	0	0	0	4
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	2	4	4	0	0	0	2	2	0	0	18

B: SMOOTHED COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	2	3	1	1	1	1	3	2	1	1	17
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	3	1	1	3	1	1	1	1	1	1	1	15
on	3	2	1	1	1	1	1	1	2	1	1	15
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	1	1	1	1	1	1	1	1	1	1	1	11
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	3	3	1	1	1	1	1	1	1	15
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	13	15	15	11	11	11	13	13	11	11	139

B: PROBABILITIES

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0153	0.0229	0.0076	0.0076	0.0076	0.0076	0.0229	0.0153	0.0076	0.0076	0.1298
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0229	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
on	0.0229	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.1145
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
hat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dog	0.0076	0.0076	0.0229	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.1145	0.0992	0.1145	0.1145	0.084	0.084	0.084	0.0992	0.0992	0.084	0.084	1.0611

the	cat	in	the	hat			
	0.0153	0.0076	0.0076	0.0076	7E-09	0.3333	2E-09
						B	0.0625

C : COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	0	2	0	0	1	1	0	0	2	1	7
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	1	0	0	0	0	1	0	0	0	0	0	2
on	0	0	0	0	0	0	0	0	0	0	0	0
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	2	0	0	0	0	0	1	0	0	0	1	4
hat	0	0	1	0	0	0	0	0	0	0	0	1
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	1	0	1	0	0	2	0	0	0	0	0	4
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	4	0	0	4	2	0	0	2	2	18

C : SMOOTHED COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	1	3	1	1	2	2	1	1	3	2	18
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	2	1	1	1	1	2	1	1	1	1	1	13
on	1	1	1	1	1	1	1	1	1	1	1	11
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	3	1	1	1	1	1	2	1	1	1	2	15
hat	1	1	2	1	1	1	1	1	1	1	1	12
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	2	1	2	1	1	3	1	1	1	1	1	15
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	11	15	11	11	15	13	11	11	13	13	139

C : PROBABILITY MATRIX

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0076	0.0229	0.0076	0.0076	0.0153	0.0153	0.0076	0.0076	0.0229	0.0153	0.1374
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0153	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
on	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0153	0.1145
hat	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
dog	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0153	0.0076	0.0153	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.1145	0.084	0.1145	0.084	0.084	0.1145	0.0992	0.084	0.084	0.0992	0.0992	1.0611

the	cat	in	the	hat	Product	P(C)	Product*P(C)
	0.0076	0.0076	0.0229	0.0153	2E-08	0.3333	7E-09
						P(d|c)P(c)/P(d)	**0.1875**

23

Suppose we have a trained first-order hidden Markov model for part-of-speech tagging with the parameters below (assume the value is 0 if the parameter is not shown in the tables).

Tag transition probabilities:

	DT	JJ	NN	VB
⟨S⟩	0.5	0.1	0.3	0.1
DT	0.0	0.5	0.5	0.0
JJ	0.0	0.4	0.6	0.0
NN	0.1	0.0	0.2	0.7
VB	0.4	0.2	0.2	0.2

Word emission probabilities:

	the	old	man	ships
DT	1.0	0.0	0.0	0.0
JJ	0.0	0.8	0.2	0.0
NN	0.0	0.2	0.4	0.4
VB	0.0	0.0	0.5	0.5

What is the predicted tag sequence for the sentence "the old man the ships," decoded using beam search Viterbi with beam size 1, and what is the predicted probability of that tag sequence? Show your work. You can leave the answer as a product of decimals.

24

Suppose we have two unigram probability distributions:

$$p(\text{cat}) = \frac{1}{2} \;\; p(\text{dog}) = \frac{1}{4}$$
$$p(\text{mat}) = \frac{1}{8} \;\; p(\text{log}) = \frac{1}{8}$$

and

$$q(\text{cat}) = \frac{1}{8} \;\; q(\text{dog}) = \frac{1}{4}$$
$$q(\text{mat}) = \frac{1}{2} \;\; q(\text{log}) = \frac{1}{8}$$

What is the cross-entropy $H(p, q)$? Show your work.

25

Suppose we have the following probabilistic context-free grammar:

NP	→	DT NBAR	1.0
NBAR	→	NN	0.4
NBAR	→	JJ NBAR	0.3
NBAR	→	NBAR NBAR	0.3
DT	→	the	1.0
JJ	→	blue	1.0
NN	→	fountain	0.3
NN	→	ink	0.3
NN	→	pen	0.4

How many possible parses are there for the phrase "the blue fountain pen ink" under this grammar? Which has the highest probability, and what is that probability? Show your work.

10. The old man the ships - viterbi beam 1

	DT	JJ	NN	VB	CHECK
<S>	0.5	0.1	0.3	0.1	1
DT	0	0.5	0.5	0	1
JJ	0	0.4	0.6	0	1
NN	0.1	0	0.2	0.7	1
VB	0.4	0.2	0.2	0.2	1

	the	old	man	ships
DT	1	0	0	0
JJ	0	0.8	0.2	0
NN	0	0.2	0.4	0.4
VB	0	0	0.5	0.5

<u>STEP - 1</u>

	the	old	man	the	ships
DT	(0.5) (1)				
JJ	(0.1) (0)				
NN	(0.1) (0)				
VB	(0.1) (0)				

	the	old	man	the	ships
DT	0.5	0			
JJ	0	0.5*0.5*0.8 = 0.2			
NN	0	0.5*0.5*0.2 = 0.05			
VB	0	0			

	the	old	man	the	ships
DT	0.5	0	0		
JJ	0	0.2	0.4*0.2*0.2 = 0.016		
NN	0	0	0.6*0.4*0.2 = 0.048		
VB	0	0	0		

	the	old	man	the	ships
DT	0.5	0	0	0.1*0.048 = 0.0048	
JJ	0	0.2	0	0	
NN	0	0.05	0.048	0	
VB	0	0	0	0	

	the	old	man	the	ships
DT	0.5	0	0	0.0048	0
JJ	0	0.2	0	0	0
NN	0	0	0.048	0	0.5*0.4*0.0048 = 0.001
VB	0	0	0	0	0

DT	JJ	NN	DT	NN

24)

$$H(p,q) = E_p\left[-\log(q)\right]$$

$$= -\left[\frac{1}{2}\log\left(\frac{1}{8}\right) + \frac{1}{4}\log\left(\frac{1}{4}\right) + \frac{1}{8}\log\left(\frac{1}{2}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right)\right]$$

$$= -\left[\frac{1}{2}(-3) + \frac{1}{4}(-2) + \frac{1}{8}(-1) + \frac{1}{8}(-3)\right]$$

$$= \frac{3}{2} + \frac{1}{2} + \frac{1}{8} + \frac{3}{8}$$

$$= 2 + \frac{1}{2} = 2.5$$

25.

Given CFG is not in CNF :

Converting the CFG into CNF gives the following rules and probabilities :

NP --> DT NBAR 1.0
NBAR --> NN 0.4
NBAR --> JJ NBAR 0.3
NBAR --> NBAR NBAR 0.3
DT --> the 1.0
JJ --> blue 1.0
NN --> fountain 0.3
NN --> ink 0.3
NN --> pen 0.4

1. There are no nullable variables
2. Removing unit variables

NP --> DT NBAR 1.0
NBAR --> NN 0.4
NBAR --> JJ NBAR 0.3
NBAR --> NBAR NBAR 0.3
DT --> the 1.0
JJ --> blue 1.0
NN --> fountain 0.3
NN --> ink 0.3
NN --> pen 0.4

NBAR --> fountain 0.4*0.3 = 0.12
NBAR --> ink 0.4*0.3 = 0.12
NBAR --> pen 0.4*0.4 = 0.16

3. Removing offending variables
4. Restructuring

NP --> DT NBAR 1.0
NBAR --> JJ NBAR 0.3
NBAR --> NBAR NBAR 0.3
DT --> the 1.0
JJ --> blue 1.0
NN --> fountain 0.3
NN --> ink 0.3
NN --> pen 0.4
NBAR --> fountain 0.12
NBAR --> ink 0.12
NBAR --> pen 0.16

	the	blue	fountain	pen	ink
the	DT 1.0				
blue		JJ 1.0			
fountain			NN 0.3, NBAR 0.12		
pen				NN 0.4, NBAR 0.16	
ink					NN 0.3, NBAR 0.12

	the	blue	fountain	pen	ink
the	DT 1.0	-			
blue		JJ 1.0	NBAR 0.3*1.0*0.12=0.036		
fountain			NN 0.3, NBAR 0.12	NBAR 0.3*0.12*0.16=0.0058	
pen				NN 0.4, NBAR 0.16	NBAR 0.3*0.16*0.12=0.0058
ink					NN 0.3, NBAR 0.12

	the	blue	fountain	pen	ink
the	DT 1.0	-	NP 1.0*1.0* 0.036 = 0.036		
blue		JJ 1.0	NBAR 0.036	NBAR 0.3*1.0*0.0058 = 0.0017, NBAR 0.3*0.036* 0.16 = 0.0017	
fountai n			NN 0.3, NBAR 0.12	NBAR 0.0058	NBAR 0.3*0.12* 0.0058 = 0.0002, NBAR 0.3*0.0058* 0.12 = 0.0002
pen				NN 0.4, NBAR 0.16	NBAR 0.0058
ink					NN 0.3, NBAR 0.12

	the	blue	fountain	pen	ink
the	DT 1.0	-	NP 0.036		
blue		JJ 1.0	NBAR 0.036	NBAR 0.0017, NBAR 0.0017	
fountain			NN 0.3, NBAR 0.12	NBAR 0.0058	NBAR 0.0002, NBAR 0.0002
pen				NN 0.4, NBAR 0.16	NBAR 0.0058
ink					NN 0.3, NBAR 0.12

	the	blue	fountain	pen	ink
the	DT 1.0	-	NP 0.036	NP 0.0017, NP 0.0017	
blue		JJ 1.0	NBAR 0.036	NBAR 0.0017, NBAR 0.0017	NBAR 0.3*0.036*0.0058 = 0.0001, NBAR 0.3*0.0017*0.12 = 0.0001 , NBAR 0.3*0.0017*0.12 = 0.0001, NBAR 0.3*1.0*0.0002 = 0.0001, NBAR 0.3*1.0*0.0002 = 0.0001
fountain			NN 0.3, NBAR 0.12	NBAR 0.0058	NBAR 0.0002, NBAR 0.0002
pen				NN 0.4, NBAR 0.16	NBAR 0.0058
ink					NN 0.3, NBAR 0.12

	the	blue	fountain	pen	ink
the	DT 1.0	-	NP 0.036	NP 0.0017, NP 0.0017	NP, NP, NP, NP, NP 0.0001
blue		JJ 1.0	NBAR 0.036	NBAR 0.0017, NBAR 0.0017	NBAR 0.0001, NBAR 0.0001 , NBAR 0.0001, NBAR 0.0001, NBAR 0.0001
fountain			NN 0.3, NBAR 0.12	NBAR 0.0058	NBAR 0.0002, NBAR 0.0002
pen				NN 0.4, NBAR 0.16	NBAR 0.0058
ink					NN 0.3, NBAR 0.12

There are five possible parses.

26

Suppose we have a training corpus of two sentences:

1.	souris	blanche	2.	souris
	white	mouse		mouse

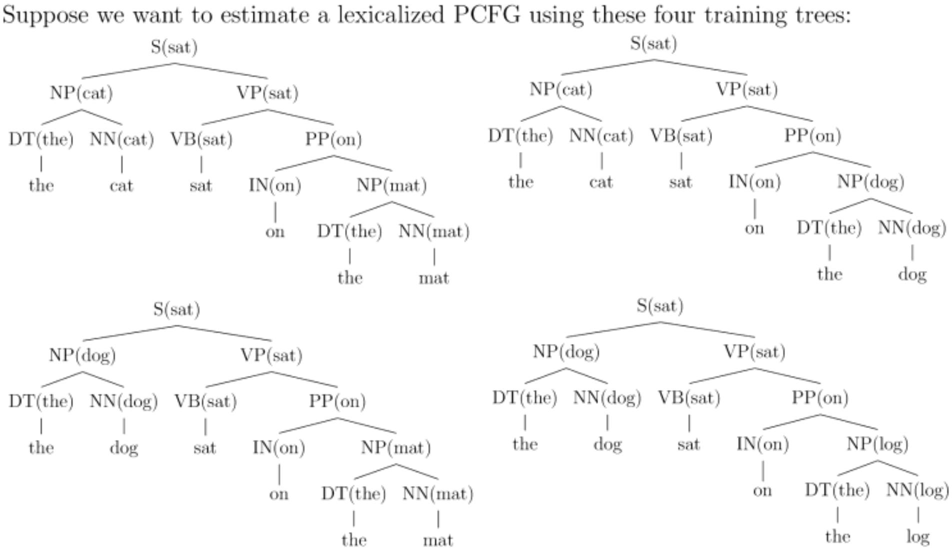
Additionally, suppose we already have a trained IBM Model 1 statistical machine translation system, and now we want to train an IBM Model 2. The parameters $q(j|i,l,m)$ are initialized uniformly, and the parameters $t(f_i|e_j)$ are initialized using the Model 1 as follows:

- $t(\text{souris}|\text{white}) = \frac{1}{2}$
- $t(\text{souris}|\text{mouse}) = \frac{3}{4}$
- $t(\text{blanche}|\text{white}) = \frac{1}{2}$
- $t(\text{blanche}|\text{mouse}) = \frac{1}{4}$

Perform one iteration of expectation maximization to update the parameters using the training corpus. What is $t(\text{blanche}|\text{white})$ at the end of this iteration? Show your work.

27

Suppose we want to estimate a lexicalized PCFG using these four training trees:



What is the probability of the rule PP(on) $\rightarrow$ IN(on) NP(dog), estimated using Charniak's method with uniform λs for smoothing? You can assume that the position of the head is fixed for each rule, so there is only one rule, not two different rules depending on head position. Show your work.

26. EM to update parameters and find t(blanche|white)

<u>INITIAL PARAMETERS:</u>

t	souris	blanche
white	1/2	1/2
mouse	3/4	1/4

We have two sentences :

For k = 1 : (m = 2, l=2)
For k = 2 : (m = 1, l=1)

Initial q's are initialized uniformly

q(j|i,l,m)
j - ENGLISH WORD
i - FRENCH WORD

q(1|1,1,1)	1		
q(1|1,2,2)	1/2	q(2|1,2,2)	1/2
q(1|2,2,2)	1/2	q(2|2,2,2)	1/2

<u>NEXT STEP:</u>

δ(k,i,j)

δ(2,1,1)	1	
δ(1,1,1)	q(1|1,2,2) t(souris|white) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] = (1/2)(1/2)/ [(1/2)(1/2) + (1/2)(3/4)]	2/5
δ(1,1,2)	q(2|1,2,2) t(souris|mouse) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] =(1/2)(3/4)/ [(1/2)(1/2) + (1/2)(3/4)]	3/5
δ(1,2,1)	q(1|2,2,2) t(blanche|white) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/2) / [(1/2)(1/2) + (1/2)(1/4)]	2/3
δ(1,2,2)	q(2|2,2,2) t(blanche|mouse) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/4) / [(1/2)(1/2) + (1/2)(1/4)]	1/3

c(white, souris)	2/5		c(white)	2/5 + 2/3		t(blanche|white)	c(white, blanche)/c(white)
c(white, blanche)	2/3		c(mouse)	3/5 + 1/3 + 1			(2/3)/ (2/5) + (2/3)
c(mouse, souris)	3/5 + 1						(10/16)
c(mouse, blanche)	1/3						

27)

$$p(PP(on) \rightarrow IN(on)\ NP(dog))$$

$$= \underbrace{p\left(PP(on) \rightarrow IN(on)\ NP \mid PP(on)\right)}_{\text{I}} \ \underbrace{p\left(dog \mid PP(on) \rightarrow IN(on)\ NP\right)}_{\text{II}}$$

Considering I : $p\left(PP(on) \rightarrow IN(on)\ NP \mid PP(on)\right)$

$$= \lambda_1\, p_{MLE}\left(PP(on) \rightarrow IN(on)\ NP \mid PP(on)\right) + \lambda_2\, p_{MLE}\left(PP \rightarrow IN\ NP \mid PP\right)$$

We have $\lambda_1 = \lambda_2 = 1/2$

$$p_{MLE}\left(PP(on) \rightarrow IN(on)\ NP \mid PP(on)\right) = 1$$

$$p_{MLE}\left(PP \rightarrow IN\ NP \mid PP\right) = 1$$

For I we have $p\left(PP(on) \rightarrow IN(on)\ NP \mid PP(on)\right) = 1$

Considering II : $p\left(dog \mid PP(on) \rightarrow IN(on)\ NP\right)$

$$= \lambda_1\, p_{MLE}\left(dog \mid PP(on) \rightarrow IN(on)\ NP\right) + \lambda_2\, p_{MLE}\left(dog \mid PP \rightarrow IN\ NP\right) + \lambda_3\, p_{MLE}\left(dog \mid NP\right)$$

We have $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$, $p_{MLE}\left(dog \mid PP(on) \rightarrow IN(on)\ NP\right) = \frac{1}{4}$

$p_{MLE}\left(dog \mid PP \rightarrow IN\,NP\right) = \frac{1}{4}$, $p_{MLE}\left(dog \mid NP\right) = \frac{3}{8}$

For II : $\frac{1}{3}\left[\frac{1}{4} + \frac{1}{4} + \frac{3}{8}\right] = \frac{1}{3}\left[\frac{4}{8} + \frac{3}{8}\right] = \frac{7}{24}$

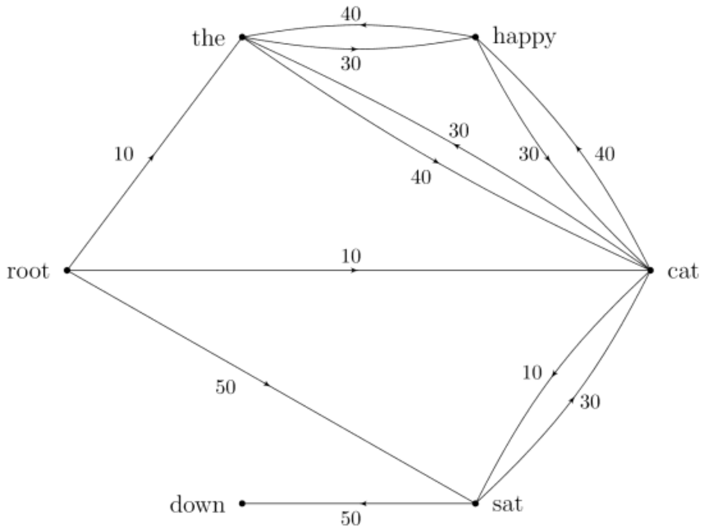
$$\therefore\ p\left(PP(on) \rightarrow IN(on)\ NP(dog)\right) = \frac{7}{24} \times 1 = \boxed{\frac{7}{24}}$$

28

Suppose we have a language model that uses linear interpolation to handle rare/out-of-vocabulary words. Is this strategy for handling rare/out-of-vocabulary words sufficient, or do we also need to use an unknown word token? Briefly explain your answer (1-2 sentences).

29

Suppose we have the following graph model of candidate dependencies for the sentence "the happy cat sat down":



You can assume that any edges not shown have score 0. What is the highest-scoring dependency parse tree based on this graph (you can give your answer using arrows (A → B) or ordered pairs (A, B)), and what is its total score? Show your work.

30

Suppose we have a training corpus of two sentences:

1.	grand	cheval	2.	grand
	big	horse		big

Now suppose we have a word-level statistical machine translation model with the parameters $q(j|i,l,m)$ and $t(f_i|e_j)$ initialized uniformly. Perform one iteration of expectation maximization to update the parameters using the training corpus. What is $t(\text{cheval}|\text{big})$ at the end of this iteration? Show your work.

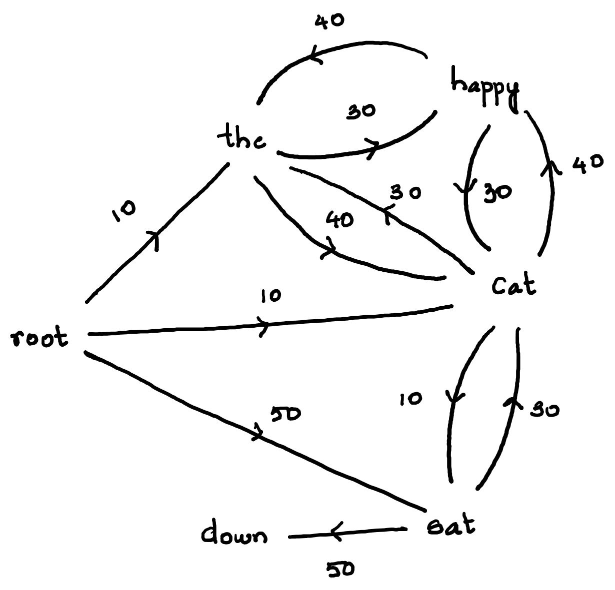
28. Linear interpolation to handle rare/out-of-vocabulary words:

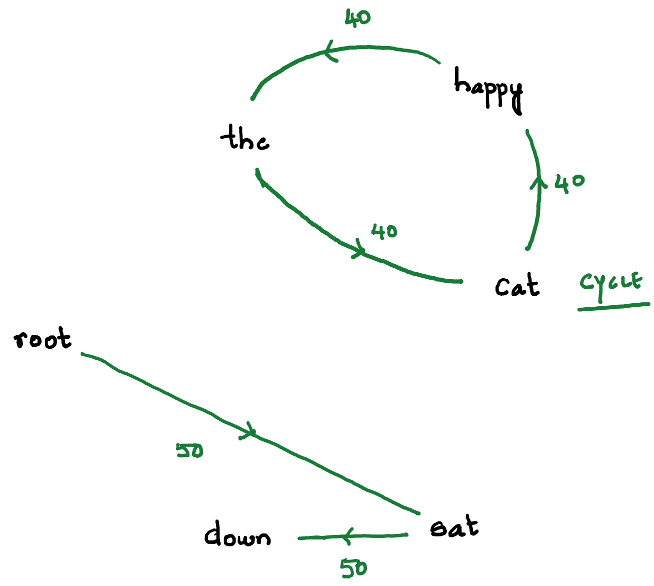
This strategy is not sufficient.
Linear interpolation extends for cases where higher n-gram probabilities are zeros.
It backs off and helps to generalize to more contexts that the model hasn't learned about.

This strategy falls short for words we haven't seen before.

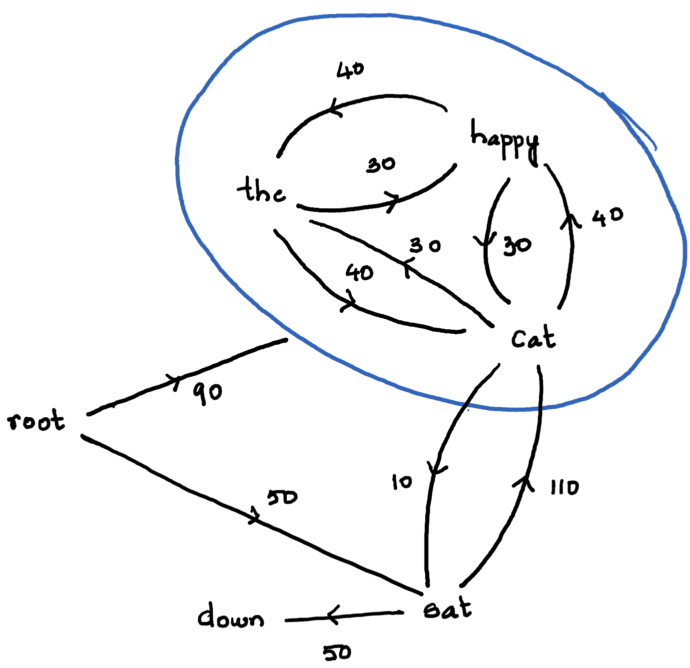
In open vocabulary system these potential unknown words are modelled using <UNK>.

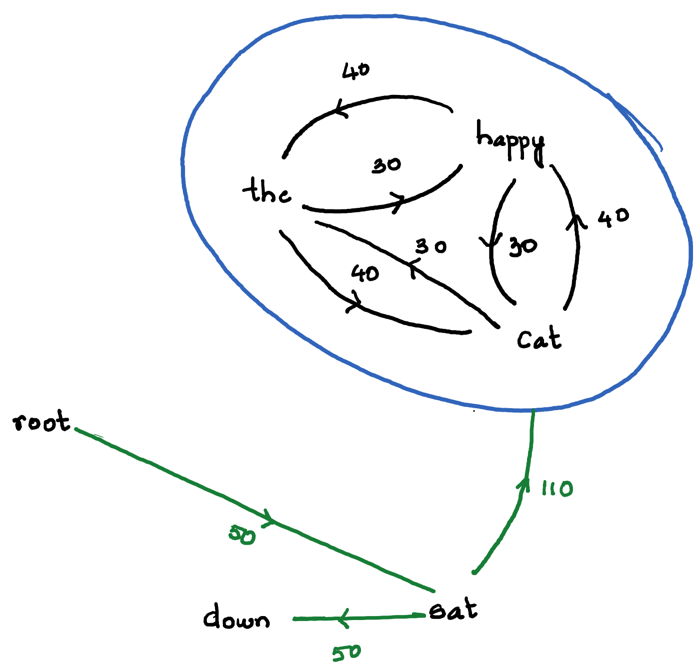
29)

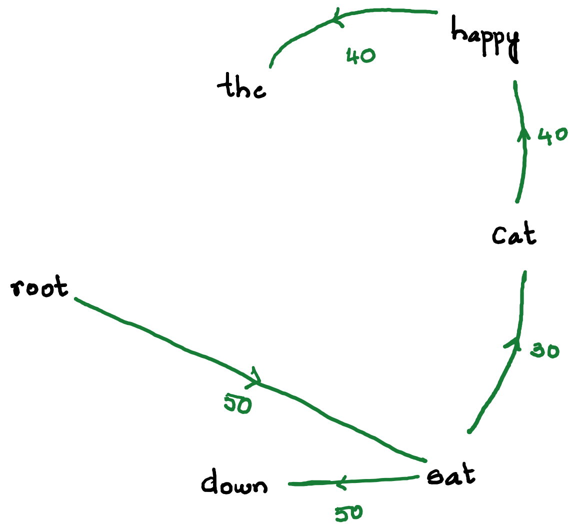




Condensed Graph :







Total score : $50 + 50 + 30 + 40 + 40$

$= 210$

30. EM to update parameters and find t(cheval|big)

INITIAL PARAMETERS: [Initialized uniformly]

t	grand	cheval
big	1/2	1/2
horse	1/2	1/2

We have two sentences :

For k = 1 : (m = 2, l=2)
For k = 2 : (m = 1, l=1)

q(j|i,l,m)
j - ENGLISH WORD
i - FRENCH WORD

q(1|1,1,1)	1		
q(1|1,2,2)	1/2	q(2|1,2,2)	1/2
q(1|2,2,2)	1/2	q(2|2,2,2)	1/2

NEXT STEP:

δ(k,i,j)

δ(2,1,1)	1	
δ(1,1,1)	q(1|1,2,2) t(grand|big) / [(q(1|1,2,2) t(grand|big) + q(2|1,2,2) t(grand|horse))] = (1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,1,2)	q(2|1,2,2) t(grand|horse) / [(q(1|1,2,2) t(grand|big) + q(2|1,2,2) t(grand|horse))] =(1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,2,1)	q(1|2,2,2) t(cheval|big) / [q(1|2,2,2) t(cheval|big) + q(2|2,2,2) t(cheval|horse))] = (1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,2,2)	q(2|2,2,2) t(cheval|horse) / [q(1|2,2,2) t(cheval|big) + q(2|2,2,2) t(cheval|horse))] =(1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2

c(big, grand)	1/2 + 1		c(big)	2		t(cheval|big)	c(big,cheval)/c(big)
c(big, cheval)	1/2		c(horse)	1			1/4
c(horse, grand)	1/2						
c(horse, cheval)	1/2						

31

Suppose you want to build a system to perform transition-based dependency parsing, and you have access to a working word sense disambiguation (WSD) system. How could you use the information provided by the WSD system to try to improve the performance of your model? Which parameter(s) of the model would you modify, and how? (2-3 sentences.)

32

Suppose we have the following probabilistic context-free grammar:

SQ	$\rightarrow$	VB S	1.0
S	$\rightarrow$	NP VP	1.0
VP	$\rightarrow$	VB NP	0.7
NP	$\rightarrow$	DT NN	1.0
DT	$\rightarrow$	the	1.0
NN	$\rightarrow$	can	0.9
NN	$\rightarrow$	see	0.1
VB	$\rightarrow$	can	0.5
VB	$\rightarrow$	see	0.5

How many possible parses are there for the sentence "can the can see the can" under this grammar? Which has the highest probability, and what is that probability? Show your work.

33

Suppose we have a trained first-order hidden Markov model for part-of-speech tagging with the parameters below (assume the value is 0 if the parameter is not shown in the tables).

Tag transition probabilities:

	DT	NN	VB	$\langle/S\rangle$
$\langle S\rangle$	0.5	0.3	0.2	0.0
DT	0.0	0.9	0.1	0.0
NN	0.1	0.2	0.3	0.4
VB	0.4	0.2	0.2	0.2

Word emission probabilities:

	the	can	see	$\langle/s\rangle$
DT	1.0	0.0	0.0	0.0
NN	0.0	0.9	0.1	0.0
VB	0.0	0.5	0.5	0.0
$\langle/S\rangle$	0.0	0.0	0.0	1.0

What is the predicted tag sequence for the sentence "can the can see $\langle/s\rangle$", decoded using greedy decoding, and what is the predicted probability of that tag sequence? Show your work. You can leave the answer as a product of decimals.

32.

	can	the	can	see	the	can
can	NN 0.9, VB 0.5					
the		DT 1.0				
can			NN 0.9, VB 0.5			
see				NN 0.1, VB 0.5		
the					DT 1.0	
can						NN 0.9, VB 0.5

	can	the	can	see	the	can
can	NN 0.9, VB 0.5	-				
the		DT 1.0	NP 1.0*1.0*0.9 = 0.9			
can			NN 0.9, VB 0.5	-		
see				NN 0.1, VB 0.5	-	
the					DT 1.0	NP 1.0*1.0*0.9 = 0.9
can						NN 0.9, VB 0.5

	can	the	can	see	the	can
can	NN 0.9, VB 0.5	-	VP 0.7*0.5*0.9 = 0.315			
the		DT 1.0	NP 0.9	-		
can			NN 0.9, VB 0.5	-	-	
see				NN 0.1, VB 0.5	-	VP 0.7*0.5*0.9 = 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	can	the	can	see	the	can
can	NN 0.9, VB 0.5	-	VP 0.315	-		
the		DT 1.0	NP 0.9	-	-	
can			NN 0.9, VB 0.5	-	-	-
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	can	the	can	see	the	can
can	NN 0.9, VB 0.5	-	VP 0.315	-	-	
the		DT 1.0	NP 0.9	-	-	S 1.0*0.9*0.315 = 0.2835
can			NN 0.9, VB 0.5	-	-	-
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	can	the	can	see	the	can
can	NN 0.9, VB 0.5	-	VP 0.315	-	-	SQ 1.0*0.5*0.2835 = 0.1418
the		DT 1.0	NP 0.9	-	-	S 0.2835
can			NN 0.9, VB 0.5	-	-	-
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	can	the	can	see	the	can
can	NN 0.9, VB 0.5	-	VP 0.315	-	-	SQ 0.1418
the		DT 1.0	NP 0.9	-	-	S 0.2835
can			NN 0.9, VB 0.5	-	-	-
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

33. Viterbi "can the can see </s>" greedy decoding

	DT	NN	VB	</S>			the	can	see	</S>
<S>	0.5	0.3	0.2	0		DT	1	0	0	0
DT	0	0.9	0.1	0		NN	0	0.9	0.1	0
NN	0.1	0.2	0.3	0.4		VB	0	0.5	0.5	0
VB	0.4	0.2	0.2	0.2		</S>	0	0	0	1

	can	the	can	see	</s>
DT	0				
NN	0.3*0.9				
VB	0.2*0.5				
</S>	0				

	can	the	can	see	</s>
DT	0	Max(0.1*1*0.27, 0.4*1*0.1) = Max(0.027, 0.04)			
NN	0.27	0			
VB	0.1	0			
</S>	0	0			

	can	the	can	see	</s>
DT	0	0.04	0		
NN	0.27	0	0.9*0.9*0.04		
VB	0.1	0	0.1*0.5*0.04		
</S>	0	0	0		

	can	the	can	see	</s>
DT	0	0.04	0	0	
NN	0.27	0	0.0324	Max(0.2*0.1*0.0324, 0.2*0.1*0.002) = Max(0.000648, 0.00004)	
VB	0.1	0	0.002	Max(0.3*0.5*0.0324, 0.2*0.5*0.002) = Max(0.00486, 0.0002)	
</S>	0	0	0	0	

	can	the	can	see	</s>
DT	0	0.04	0	0	0
NN	0.27	0	0.0324	0.000648	0
VB	0.1	0	0.002	0.00486	0
</S>	0	0	0	0	Max(0.4*0.000648, 0.2*0.00486)= Max(0.0002592, 0.000972)

	can	the	can	see	</s>
DT	0	0.04	0	0	0
NN	0.27	0	0.0324	0.000648	0
VB	0.1	0	0.002	0.00486	0
</S>	0	0	0	0	0.000972

VB	DT	NN	VB	</S>

34

Which of the following underlined phrases is a constituent? Choose all that apply.

A These black cats detest those <u>green peas</u>.
B These <u>black cats</u> detest those green peas.
C Put it <u>over on the table</u>.
D Put it over <u>on the table</u>.

35

Suppose we train a hidden Markov model to do part-of-speech tagging using the following training sequences:

- the/DT cat/NN is/VB in/PP the/DT box/NN
- the/DT cat/NN sat/VB on/PP the/DT mat/NN
- the/DT black/JJ cat/NN sat/VB

We use a start tag $\langle S \rangle$, but no end tag, bigram tag transitions, and we smooth the transition probabilities using interpolation with $\lambda_1 = \lambda_2 = \frac{1}{2}$. What is the predicted tag sequence for "the mat is black" under this model, and what is its probability? Show your work. You can leave your answer as a product of fractions.

36

"Projectivize" the dependency parse tree of the sentence "I saw a dog on my walk that was wearing a hat" using Nivre and Nilsson's method. What arc(s) would you need to lift, and what new arc(s) would you replace them with? You can give your answer using arrows ($A \rightarrow B$) or ordered pairs (A, B).

37

Which of the following underlined phrases is a constituent? Choose all that apply.

A Ali Baba returned from this travels <u>wiser than</u> before.
B The cat walked across the porch <u>with a</u> confident air.
C They arrived at the concert <u>more quickly than they expected</u>.

34. Constituent

A These black cats detest those green peas.
B These black cats detest those green peas.
C Put it over on the table.
D Put it over on the table.

A. Not a constituent
B. Not a constituent
C. Constituent
D. Constituent

35.

<S>	DT	NN	VB	PP	DT	NN
<S>	DT	NN	VB	PP	DT	NN
<S>	DT	JJ	NN	VB		

TRANSITION PROBABILITIES

	DT	NN	VB	PP	JJ
<S>	1	0	0	0	0
DT	0	0.8	0	0	0.2
NN	0	0	1	0	0
VB	0	0	0	1	0
PP	1	0	0	0	0
JJ	0	1	0	0	0

UNIGRAM TRANSITION PROBABILITIES

	c	P
<S>	3	0.157895
DT	5	0.263158
NN	5	0.263158
VB	3	0.157895
PP	2	0.105263
JJ	1	0.052632
	19	

SMOOTHED TRANSITION PROBABILITIES

	DT	NN	VB	PP	JJ
<S>	0.578947	0.078947	0.078947	0.078947	0.078947
DT	0.131579	0.531579	0.131579	0.131579	0.231579
NN	0.131579	0.131579	0.631579	0.131579	0.131579
VB	0.078947	0.078947	0.078947	0.578947	0.078947
PP	0.552632	0.052632	0.052632	0.052632	0.052632
JJ	0.026316	0.526316	0.026316	0.026316	0.026316

EMISSION PROBABILITIES

	the	cat	is	in	box	sat	on	mat	black
DT	1	0	0	0	0	0	0	0	0
NN	0	0.6	0	0	0.2	0	0	0.2	0
VB	0	0	0.5	0	0	0.5	0	0	0
PP	0	0	0	0.5	0	0	0.5	0	0
JJ	0	0	0	0	0	0	0	0	1

POS TAGGING

	the	mat	is	black
DT	0.578			
NN	0			
VB	0			
PP	0			
JJ	0			

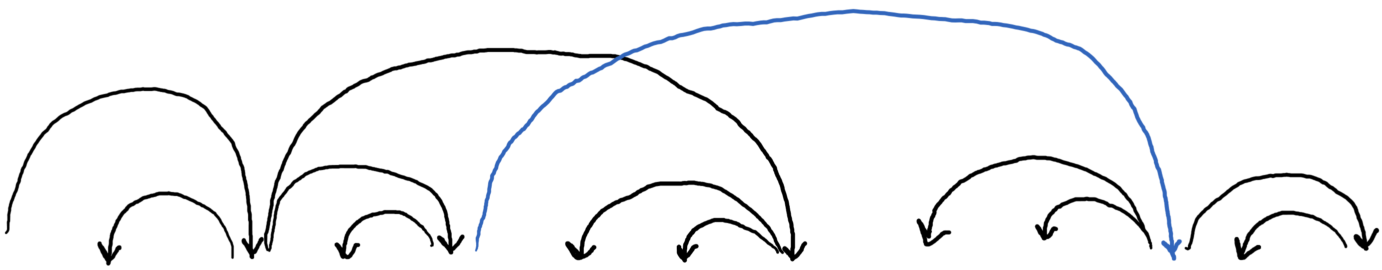
	the	mat	is	black
DT	0.578	0		
NN	0	0.578*0.5315*0.2 = 0.0614		
VB	0	0		
PP	0	0		
JJ	0	0		

	the	mat	is	black
DT	0.578	0	0	
NN	0	0.0614	0	
VB	0	0	0.0614*0.6315*0.5 = 0.0194	
PP	0	0	0	
JJ	0	0	0	

	the	mat	is	black
DT	0.578	0	0	0
NN	0	0.0614	0	0
VB	0	0	0.0194	0
PP	0	0	0	0
JJ	0	0	0	0.0194*0.0789*1 = 0.0015

DT NN VB JJ

36. Projectivize



root	I	saw	a	dog	on	my	walk	that	was	wearing	a	hat

root	I	saw	a	dog	on	my	walk	that	was	wearing	a	hat

37. Constituent

A Ali Baba returned from this travels <u>wiser than</u> before.
B The cat walked across the porch <u>with a</u> confident air.
C They arrived at the concert <u>more quickly than they expected.</u>

A. Not a constituent
B. Not a constituent
C. Constituent

38

Suppose we train a language model on the following training sequences:

- the man saw the dog with the telescope
- the dog saw the man in the park
- the man with the telescope saw the park

We use linear interpolation between a bigram model and a unigram model, with $\lambda_1 = \lambda_2 = 0.5$. The bigram model is smoothed using Kneser-Ney smoothing with $\beta = 1$, and the unigram model is not smoothed. We do not use any start or end tokens. What is $p(\text{saw}|\text{telescope})$ under this model? Show your work. You can leave the answer as a sum of fractions.

39

Suppose you want to build a system to perform graph-based dependency parsing, and you have access to a working word sense disambiguation (WSD) system. How could you use the information provided by the WSD system to try to improve the performance of your model? Which parameter(s) of the model would you modify, and how? (2-3 sentences.)

38.

Suppose we train a language model on the following training sequences:
the man saw the dog with the telescope
the dog saw the man in the park
the man with the telescope saw the park

p(saw|telescope)

Linear interpolation :

1. Bigram + Knesser-Ney smoothing (β=1)
2. Unigram Model

c(u,v)	the	man	saw	dog	with	telescope	in	park	SUM	c(u,v)>0
the	0	3	0	2	0	2	0	2	9	4
man	0	0	1	0	1	0	1	0	3	3
saw	3	0	0	0	0	0	0	0	3	1
dog	0	0	1	0	1	0	0	0	2	2
with	2	0	0	0	0	0	0	0	2	1
telescope	0	0	1	0	0	0	0	0	1	1
in	1	0	0	0	0	0	0	0	1	1
park	0	0	0	0	0	0	0	0	0	0
									21	13

<u>DISCOUNTED COUNTS</u>

										$\lambda(w_{i-1})$
c(u,v)	the	man	saw	dog	with	telescope	in	park		
the	0	2	0	1	0	1	0	1		0.44444
man	0	0	0	0	0	0	0	0		1
saw	2	0	0	0	0	0	0	0		0.33333
dog	0	0	0	0	0	0	0	0		1
with	1	0	0	0	0	0	0	0		0.5
telescope	0	0	0	0	0	0	0	0		**0.5**
in	0	0	0	0	0	0	0	0		1
park	0	0	0	0	0	0	0	0		0

P(saw|telescope)

λ(telescope) = 0.5

	c(u)	p
the	9	0.38
man	3	0.13
saw	3	0.13
dog	2	0.08
with	2	0.08
telescope	2	0.08
in	1	0.04
park	2	0.08
	24	1

P(saw|telescope)

= 0.5*[Ps(saw|telescope)] + 0.5*P(saw)
= 0.5*[c_discounted(telescope, saw)/c() + λ(telescope) * P_cont(saw)] + 0.5*(3/24)

c_discounted(telescope, saw) = 1-1 = 0

λ(telescope) = 1
P_cont(saw) = 1/13

= 0.5* [1/13] + 0.5 * (3/24)

40

Suppose we have the following context-free grammar:

$$
\begin{array}{lcl}
\text{S} & \rightarrow & \text{NP VP} \\
\text{VP} & \rightarrow & \text{VP PP} \\
\text{VP} & \rightarrow & \text{VB NP} \\
\text{VP} & \rightarrow & \text{VB} \\
\text{NP} & \rightarrow & \text{DT NOM} \\
\text{NOM} & \rightarrow & \text{JJ NOM} \\
\text{NOM} & \rightarrow & \text{NN} \\
\text{PP} & \rightarrow & \text{IN NP} \\
\text{DT} & \rightarrow & \text{the} \\
\text{DT} & \rightarrow & \text{in} \\
\text{IN} & \rightarrow & \text{under} \\
\text{JJ} & \rightarrow & \text{big} \\
\text{JJ} & \rightarrow & \text{white} \\
\text{JJ} & \rightarrow & \text{yellow} \\
\text{NN} & \rightarrow & \text{dog} \\
\text{NN} & \rightarrow & \text{hat} \\
\text{NN} & \rightarrow & \text{house} \\
\text{VB} & \rightarrow & \text{ran}
\end{array}
$$

What new rule(s) would you have to add to the grammar to be able to parse the sentence "the dog in the white hat ran under the big yellow house"? If you need new symbols, use X, Y, and Z.

41

"Projectivize" the dependency parse tree of the sentence "What is that dog chewing on over there?" using Nivre and Nilsson's method. What arc(s) would you need to lift, and what new arc(s) would you replace them with? You can give your answer using arrows (A $\rightarrow$ B) or ordered pairs (A, B).

42

Suppose we have a training corpus of two sentences:

1.	grand	cheval	2.	grand
	big	horse		big

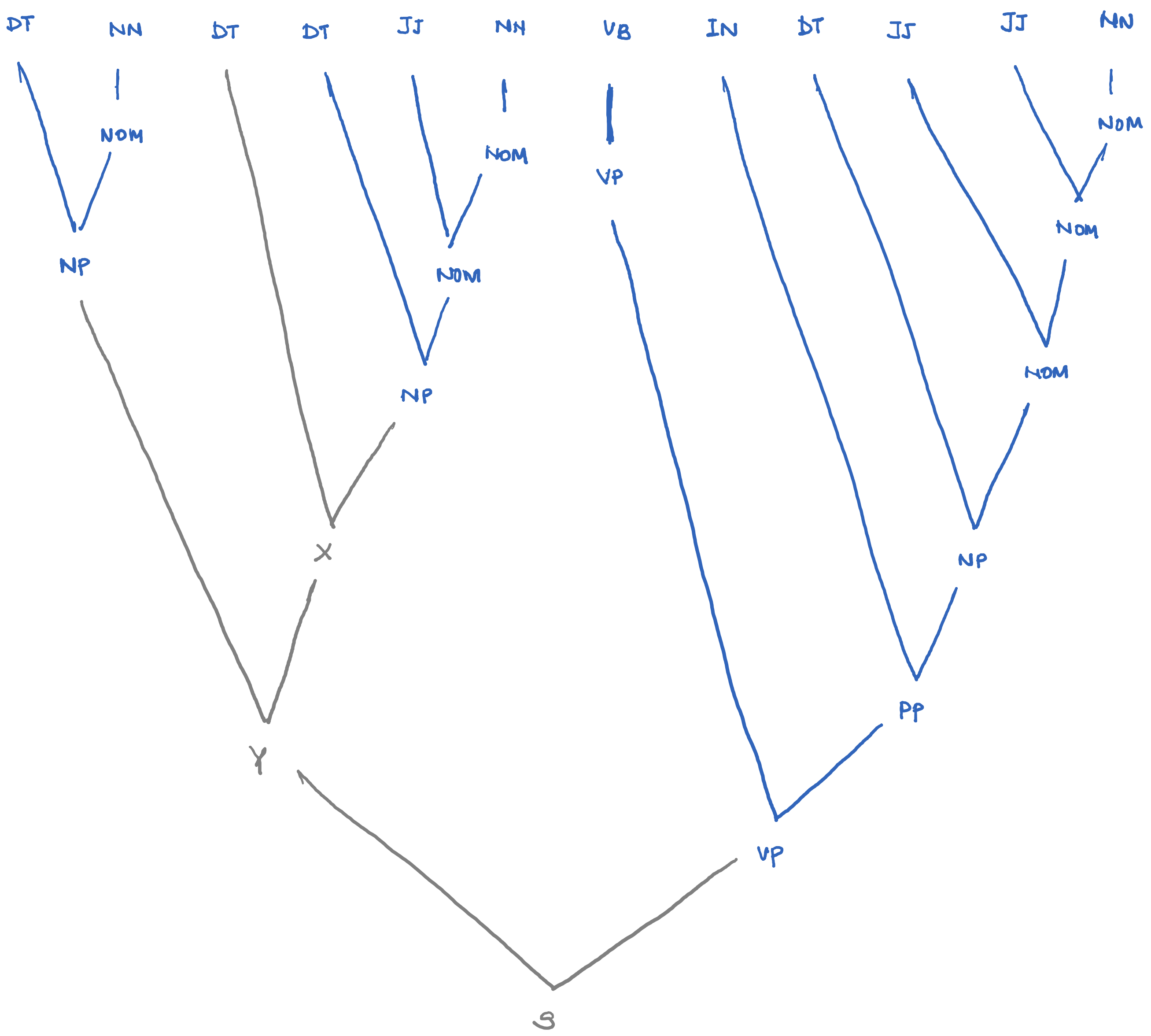
Now suppose we have a word-level statistical machine translation model with the parameters $q(j|i,l,m)$ and $t(f_i|e_j)$ initialized uniformly. Perform one iteration of expectation maximization to update the parameters using the training corpus. What is $t(\text{grand}|\text{big})$ at the end of this iteration? Show your work.

40.

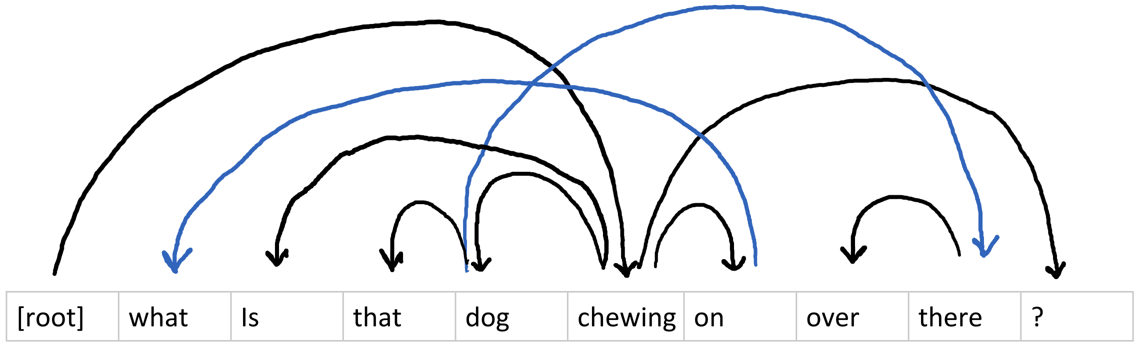
New rules to be added to parse :

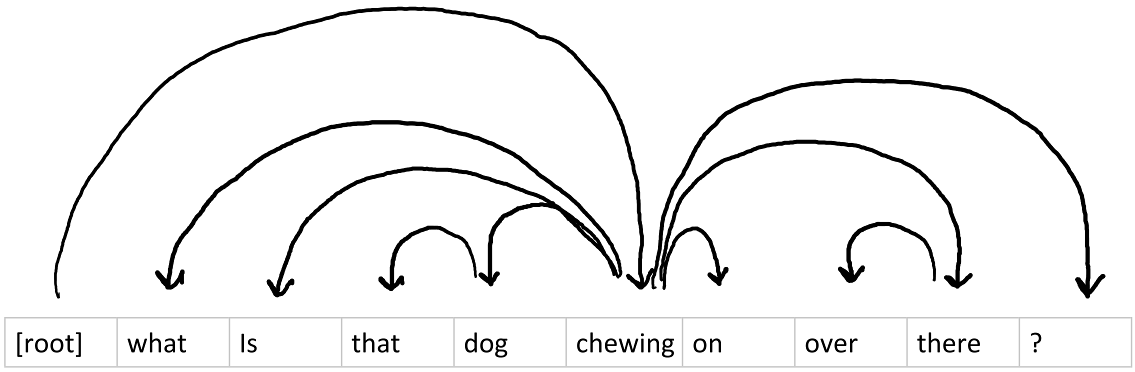
X --> DT NP
Y --> NP X
S --> Y VP

the	dog	in	the	white	hat	ran	under	the	big	yellow	house



13


[root]
what
Is
that
dog
chewing
on
over
there
?


[root]
what
Is
that
dog
chewing
on
over
there
?

42. EM to update parameters and find t(grand|big)

INITIAL PARAMETERS: [Initialized uniformly]

t	grand	cheval
big	1/2	1/2
horse	1/2	1/2

We have two sentences :

For k = 1 : (m = 2, l=2)
For k = 2 : (m = 1, l=1)

q(j|i,l,m)
j - ENGLISH WORD
i - FRENCH WORD

q(1|1,1,1)	1		
q(1|1,2,2)	1/2	q(2|1,2,2)	1/2
q(1|2,2,2)	1/2	q(2|2,2,2)	1/2

NEXT STEP:

δ(k,i,j)

δ(2,1,1)	1	
δ(1,1,1)	q(1|1,2,2) t(grand|big) / [(q(1|1,2,2) t(grand|big) + q(2|1,2,2) t(grand|horse))] = (1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,1,2)	q(2|1,2,2) t(grand|horse) / [(q(1|1,2,2) t(grand|big) + q(2|1,2,2) t(grand|horse))] =(1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,2,1)	q(1|2,2,2) t(cheval|big) / [q(1|2,2,2) t(cheval|big) + q(2|2,2,2) t(cheval|horse))] = (1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2
δ(1,2,2)	q(2|2,2,2) t(cheval|horse) / [q(1|2,2,2) t(cheval|big) + q(2|2,2,2) t(cheval|horse))] =(1/2)(1/2)/ [(1/2)(1/2) +(1/2)(1/2)]	1/2

c(big, grand)	1/2 + 1		c(big)	2		t(grand|big)	c(big, grand)/c(big)
c(big, cheval)	1/2		c(horse)	1			3/4
c(horse, grand)	1/2						
c(horse, cheval)	1/2						

43

Suppose we have the following context-free grammar:

S	$\rightarrow$	NP VP
VP	$\rightarrow$	VB NP
NP	$\rightarrow$	NP and NP
NP	$\rightarrow$	JJ NP
NP	$\rightarrow$	NN
JJ	$\rightarrow$	black
NN	$\rightarrow$	I
NN	$\rightarrow$	cats
NN	$\rightarrow$	dogs
VB	$\rightarrow$	like

What new rule(s) would you have to add to the grammar to be able to parse the sentences "I think I like black cats" and "I heard cats like dogs"? If you need new symbols, use X, Y, and Z.

44

Suppose we have the following probabilistic context-free grammar:

S	$\rightarrow$	NP VP	1.0
VP	$\rightarrow$	VB VP	0.3
VP	$\rightarrow$	VB NP	0.7
NP	$\rightarrow$	DT NN	1.0
DT	$\rightarrow$	the	1.0
NN	$\rightarrow$	can	0.9
NN	$\rightarrow$	see	0.1
VB	$\rightarrow$	can	0.5
VB	$\rightarrow$	see	0.5

How many possible parses are there for the sentence "the can can see the can" under this grammar? Which has the highest probability, and what is that probability? Show your work.

43.

New rules to be added in order to produce

1. I think I like black cats
2. I heard cats like dogs

X --> think
Y --> heard

VP --> X S
VP --> Y S

44.

	the	can	can	see	the	can
the	DT 1.0					
can		NN 0.9, VB 0.5				
can			NN 0.9, VB 0.5			
see				NN 0.1, VB 0.5		
the					DT 1.0	
can						NN 0.9, VB 0.5

	the	can	can	see	the	can
the	DT 1.0	NP 0.9				
can		NN 0.9, VB 0.5	-			
can			NN 0.9, VB 0.5	-		
see				NN 0.1, VB 0.5	-	
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	the	can	can	see	the	can
the	DT 1.0	NP 0.9	-			
can		NN 0.9, VB 0.5	-	-		
can			NN 0.9, VB 0.5	-	-	
see				NN 0.1, VB 0.5	-	VP 0.7*0.5*0.9
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	the	can	can	see	the	can
the	DT 1.0	NP 0.9	-	-		
can		NN 0.9, VB 0.5	-	-	-	
can			NN 0.9, VB 0.5	-	-	VP 0.3*0.315*0.5
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

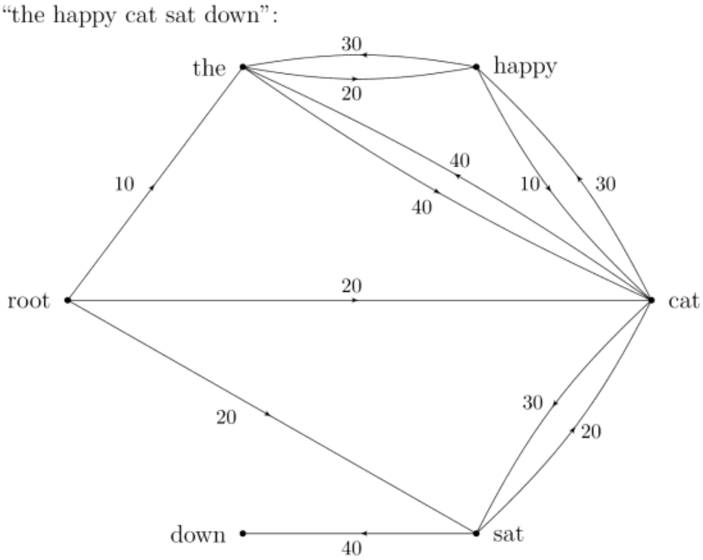
	the	can	can	see	the	can
the	DT 1.0	NP 0.9	-	-	-	
can		NN 0.9, VB 0.5	-	-	-	VP 0.3*0.04725*0.5
can			NN 0.9, VB 0.5	-	-	VP 0.04725
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	the	can	can	see	the	can
the	DT 1.0	NP 0.9	-	-	-	S 1*0.04725*0.9
can		NN 0.9, VB 0.5	-	-	-	VP 0.0070875
can			NN 0.9, VB 0.5	-	-	VP 0.04725
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

	the	can	can	see	the	can
the	DT 1.0	NP 0.9	-	-	-	S 1*0.04725*0.9
can		NN 0.9, VB 0.5	-	-	-	VP 0.0070875
can			NN 0.9, VB 0.5	-	-	VP 0.04725
see				NN 0.1, VB 0.5	-	VP 0.315
the					DT 1.0	NP 0.9
can						NN 0.9, VB 0.5

45

Suppose we have the following graph model of candidate dependencies for the sentence "the happy cat sat down":



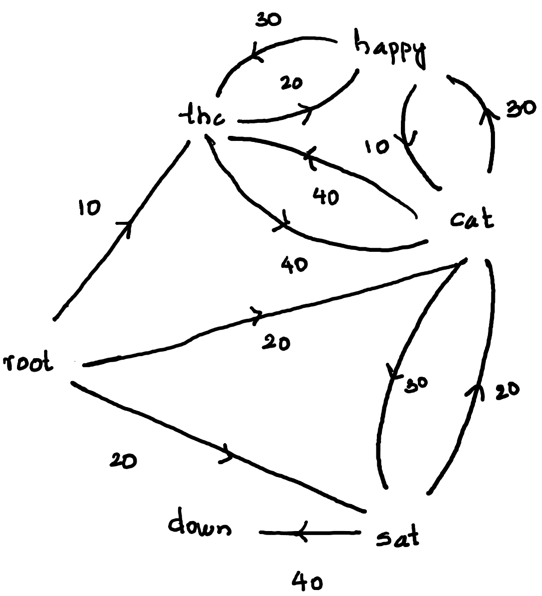
You can assume that any edges not shown have score 0. What is the highest-scoring dependency parse tree based on this graph (you can give your answer using arrows (A → B) or ordered pairs (A, B)), and what is its total score? Show your work.

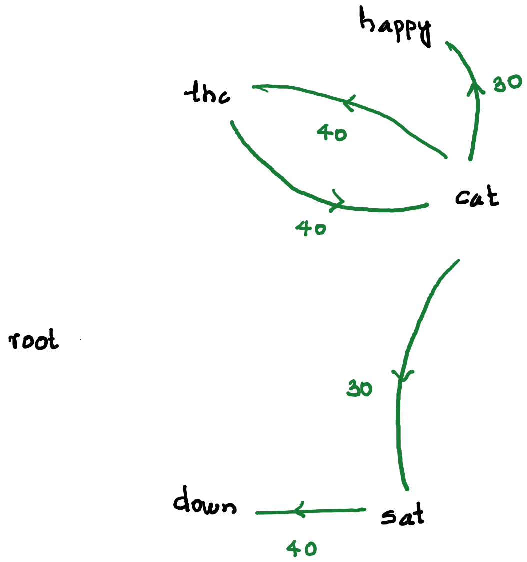
46

Which of the following underlined phrases is a constituent? Choose all that apply.

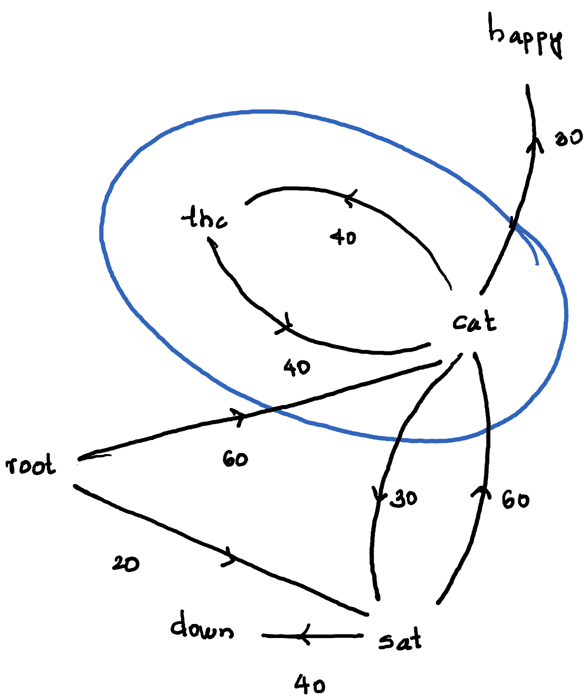
A The cat walked across the porch <u>with a confident air</u>.
B They arrived at the concert <u>more quickly than</u> they expected.
C I am <u>very fond of my nephew</u>.

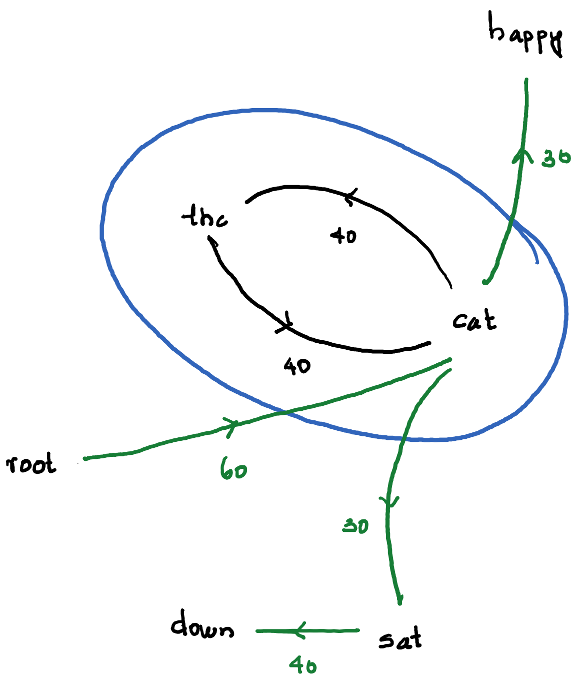
45)

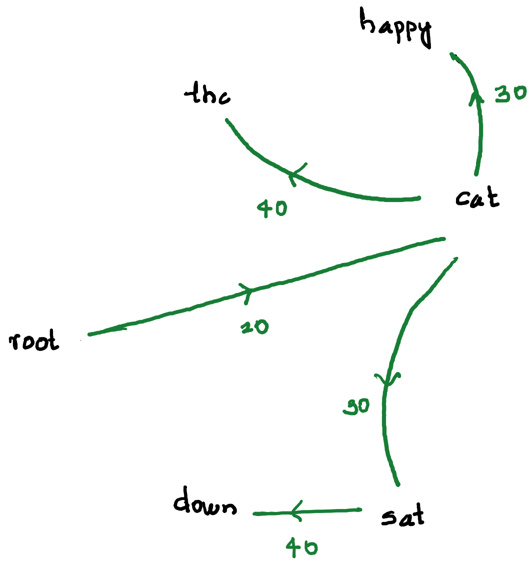




CONDENSED GRAPH







Total score : $40+30+30+20+40$

$= 160$

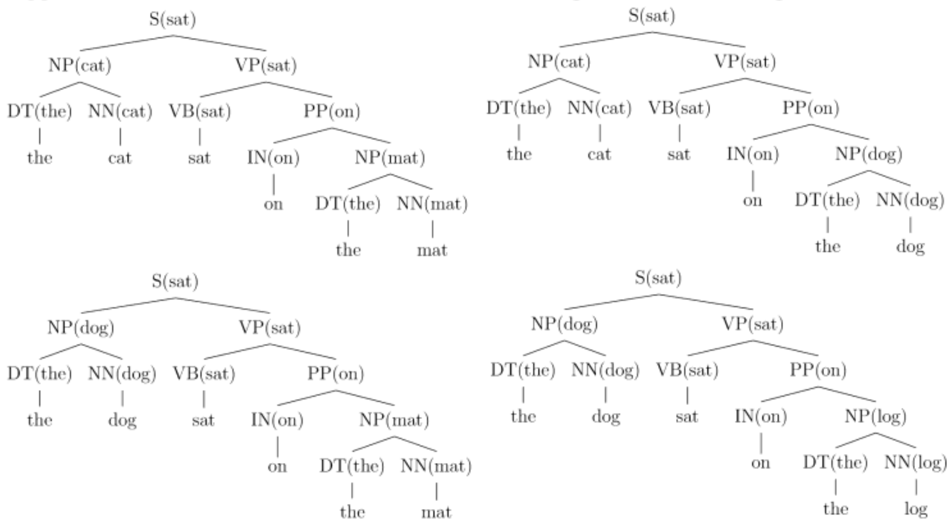
46.

A The cat walked across the porch <u>with a confident air.</u>
B They arrived at the concert <u>more quickly than</u> they expected.
C I am <u>very fond of my nephew.</u>

A. Constituent
B. Not a constituent
C. Constituent

47

Suppose we want to estimate a lexicalized PCFG using these four training trees:



What is the probability of the rule PP(on) $\rightarrow$ IN(on) NP(cat), estimated using Charniak's method with uniform λs for smoothing? You can assume that the position of the head is fixed for each rule, so there is only one rule, not two different rules depending on head position. Show your work.

48

Suppose we have a dataset of restaurant reviews, and we want to perform binary sentiment classification. However, the restaurant corpus does not come with gold standard labels. We have another dataset consisting of movie reivews that does come with gold standard labels. How could you use the movie corpus, along with bootstrapping, to develop a system to perform sentiment classification on the restaurant corpus? Describe how you would design and train the system (3-4 senctences).

47) $p(PP(on) \rightarrow IN(on)\ NP(cat))$

$$= \underbrace{p(PP(on) \rightarrow IN(on)\ NP \mid PP(on))}_{\text{I}}\ \underbrace{p(cat \mid PP(on) \rightarrow IN(on)\ NP)}_{\text{II}}$$

Considering (I): $p(PP(on) \rightarrow IN(on)\ NP \mid PP(on))$

$$= \lambda_1\, p_{MLE}(PP(on) \rightarrow IN(on)\ NP \mid PP(on)) + \lambda_2\, p_{MLE}(PP \rightarrow IN\ NP \mid PP)$$

We have $\lambda_1 = \lambda_2 = 1/2$

$$p_{MLE}(PP(on) \rightarrow IN(on)\ NP \mid PP(on)) = 1$$

$$p_{MLE}(PP \rightarrow IN\ NP \mid PP) = 1$$

For (I) we have $p(PP(on) \rightarrow IN(on)\ NP \mid PP(on)) = 1$

Considering (II): $p(cat \mid PP(on) \rightarrow IN(on)\ NP)$

$$= \lambda_1\, p_{MLE}(cat \mid PP(on) \rightarrow IN(on)\ NP) + \lambda_2\, p_{MLE}(cat \mid PP \rightarrow IN\ NP) + \lambda_3\, p_{MLE}(cat \mid NP)$$

We have $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$, $p_{MLE}(cat \mid PP(on) \rightarrow IN(on)\ NP) = 0$

$$p_{MLE}(cat \mid PP \rightarrow IN\ NP) = 0, \quad p_{MLE}(cat \mid NP) = \frac{1}{4}$$

$$\Rightarrow p(cat \mid PP(on) \rightarrow IN(on)\ NP) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$$

$$\therefore p(PP(on) \rightarrow IN(on)\ NP(cat)) = \boxed{\frac{1}{12}}$$

49

Suppose we have a training corpus of two sentences:

1.	souris	blanche	2.	souris
	white	mouse		mouse

Additionally, suppose we already have a trained IBM Model 1 statistical machine translation system, and now we want to train an IBM Model 2. The parameters $q(j|i,l,m)$ are initialized uniformly, and the parameters $t(f_i|e_j)$ are initialized using the Model 1 as follows:

- $t(\text{souris}|\text{white}) = \frac{1}{2}$
- $t(\text{souris}|\text{mouse}) = \frac{3}{4}$
- $t(\text{blanche}|\text{white}) = \frac{1}{2}$
- $t(\text{blanche}|\text{mouse}) = \frac{1}{4}$

Perform one iteration of expectation maximization to update the parameters using the training corpus. What is $t(\text{souris}|\text{mouse})$ at the end of this iteration? Show your work.

50

Suppose you want to build a system to perform constituent parsing using a context-free grammar, and you have access to a working word sense disambiguation (WSD) system. How could you use the information provided by the WSD system to try to improve the performance of your model? Which parameter(s) of the model would you modify, and how? (2-3 sentences.)

51

Suppose we train a Naive Bayes classifier on the following training sequences (the letter after the comma is the class label):

- the cat sat on the mat, A
- the cat sat in the hat, A
- the dog sat on the log, B
- the dog sat on the cat, B
- the fish sat in the dish, C
- the fish in the hat sat, C

Our classifier uses skipgram count features, no start or end tokens, and Laplace smoothing with $\delta = 1$. Given the test sentence "the cat in the hat," what is the predicted probability of this sentence belonging to class A under this model? Show your work. You can leave the answer as a product of fractions.

49.
EM to update parameters and find t(souris|mouse)

INITIAL PARAMETERS:

t	souris	blanche
white	1/2	1/2
mouse	3/4	1/4

We have two sentences :

For k = 1 : (m = 2, l=2)
For k = 2 : (m = 1, l=1)

Initial q's are initialized uniformly

q(j|i,l,m)
j - ENGLISH WORD
i - FRENCH WORD

q(1|1,1,1)	1		
q(1|1,2,2)	1/2	q(2|1,2,2)	1/2
q(1|2,2,2)	1/2	q(2|2,2,2)	1/2

NEXT STEP:

δ(k,i,j)

δ(2,1,1)	1	
δ(1,1,1)	q(1|1,2,2) t(souris|white) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] = (1/2)(1/2)/ [(1/2)(1/2) + (1/2)(3/4)]	2/5
δ(1,1,2)	q(2|1,2,2) t(souris|mouse) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] =(1/2)(3/4)/ [(1/2)(1/2) + (1/2)(3/4)]	3/5
δ(1,2,1)	q(1|2,2,2) t(blanche|white) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/2) / [(1/2)(1/2) + (1/2)(1/4)]	2/3
δ(1,2,2)	q(2|2,2,2) t(blanche|mouse) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/4) / [(1/2)(1/2) + (1/2)(1/4)]	1/3

c(white, souris)	2/5		c(white)	2/5 + 2/3		t(souris|mouse)	c(mouse,souris)/c(mouse)
c(white, blanche)	2/3		c(mouse)	3/5 + 1/3 + 1			(3/5 + 1)/[3/5 + 1/3 + 1]
c(mouse, souris)	3/5 + 1						(8/5) / [29/15]
c(mouse, blanche)	1/3						24/29

51. Naive Bayes

1. the cat sat on the mat, A
2. the cat sat in the hat, A
3. the dog sat on the log, B
4. the dog sat on the cat, B
5. the fish sat in the dish, C
6. the fish in the hat sat, C

Skipgram features
Laplace smoothing

Skipgram : We also include counts of skipgrams

A :
(the, sat) - 2, (cat, on) - 1, (sat, the) - 2, (on, mat) - 1, (cat, in) - 1, (in, hat) - 1

B :
(the,sat) - 2, (dog, on) - 2, (sat, the) - 2, (on, log) - 1, (on, cat) - 1

C:
(the, sat) - 2, (the, in) - 1, (fish, in) - 1, (fish, the) - 1, (sat, the) - 1, (in, hat) - 1, (in, dish) - 1

A : COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	2	2	0	1	0	1	0	0	0	0	6
cat	0	0	2	1	0	1	0	0	0	0	0	4
sat	2	0	0	1	0	1	0	0	0	0	0	4
on	1	0	0	0	1	0	0	0	0	0	0	2
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	1	0	0	0	0	0	1	0	0	0	0	2
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	2	4	2	2	2	2	0	0	0	0	18

A : SMOOTHED COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	3	3	1	2	1	2	1	1	1	1	17
cat	1	1	3	2	1	2	1	1	1	1	1	15
sat	3	1	1	2	1	2	1	1	1	1	1	15
on	2	1	1	1	2	1	1	1	1	1	1	13
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	2	1	1	1	1	1	2	1	1	1	1	13
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	13	15	13	13	13	13	11	11	11	11	139

A : PROBABILITY MATRIX

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.00763	0.0229	0.0229	0.0076	0.0153	0.0076	0.0153	0.00763359	0.0076	0.007633588	0.0076	0.1298
cat	0.00763	0.0076	0.0229	0.0153	0.0076	0.0153	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.1145
sat	0.0229	0.0076	0.0076	0.0153	0.0076	0.0153	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.1145
on	0.01527	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.0992
mat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
in	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.00763359	0.0076	0.007633588	0.0076	0.0992
hat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
dog	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
log	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
fish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
dish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.00763359	0.0076	0.007633588	0.0076	0.084
	0.1145	0.0992	0.1145	0.0992	0.0992	0.0992	0.0992	0.08396947	0.084	0.083969466	0.084	1.0611

the	cat	in	the	hat						
	0.0229	0.0153	0.0153	0.0153	8.1494E-08	0.3333	2.71647E-08		SUM	3.62196E-08
						A	0.75			

B: COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	1	2	0	0	0	0	2	1	0	0	6
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	2	0	0	2	0	0	0	0	0	0	0	4
on	2	1	0	0	0	0	0	0	1	0	0	4
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	0	0	0	0	0	0	0	0	0	0	0	0
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	2	2	0	0	0	0	0	0	0	4
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	2	4	4	0	0	0	2	2	0	0	18

B: SMOOTHED COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	2	3	1	1	1	1	3	2	1	1	17
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	3	1	1	3	1	1	1	1	1	1	1	15
on	3	2	1	1	1	1	1	1	2	1	1	15
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	1	1	1	1	1	1	1	1	1	1	1	11
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	3	3	1	1	1	1	1	1	1	15
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	13	15	15	11	11	11	13	13	11	11	139

B: PROBABILITIES

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0153	0.0229	0.0076	0.0076	0.0076	0.0076	0.0229	0.0153	0.0076	0.0076	0.1298
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0229	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
on	0.0229	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.1145
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
hat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dog	0.0076	0.0076	0.0229	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.1145	0.0992	0.1145	0.1145	0.084	0.084	0.084	0.0992	0.0992	0.084	0.084	1.0611

the	cat	in	the	hat			
	0.0153	0.0076	0.0076	0.0076	7E-09	0.3333	2E-09
						B	0.0625

C : COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	0	2	0	0	1	1	0	0	2	1	7
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	1	0	0	0	0	1	0	0	0	0	0	2
on	0	0	0	0	0	0	0	0	0	0	0	0
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	2	0	0	0	0	0	1	0	0	0	1	4
hat	0	0	1	0	0	0	0	0	0	0	0	1
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	1	0	1	0	0	2	0	0	0	0	0	4
dish	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	4	0	0	4	2	0	0	2	2	18

C : SMOOTHED COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	1	3	1	1	2	2	1	1	3	2	18
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	2	1	1	1	1	2	1	1	1	1	1	13
on	1	1	1	1	1	1	1	1	1	1	1	11
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	3	1	1	1	1	1	2	1	1	1	2	15
hat	1	1	2	1	1	1	1	1	1	1	1	12
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	2	1	2	1	1	3	1	1	1	1	1	15
dish	1	1	1	1	1	1	1	1	1	1	1	11
	15	11	15	11	11	15	13	11	11	13	13	139

C : PROBABILITY MATRIX

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0076	0.0229	0.0076	0.0076	0.0153	0.0153	0.0076	0.0076	0.0229	0.0153	0.1374
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0153	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
on	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0153	0.1145
hat	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
dog	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0153	0.0076	0.0153	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.1145
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.1145	0.084	0.1145	0.084	0.084	0.1145	0.0992	0.084	0.084	0.0992	0.0992	1.0611

the	cat	in	the	hat	Product	P(C)	Product*P(C)
	0.0076	0.0076	0.0229	0.0153	2E-08	0.3333	7E-09
						P(d|c)P(c)/P(d)	**0.1875**

52

Which of the following underlined phrases is a constituent? Choose all that apply.

A These black cats detest those green peas.
B These black cats detest those green peas.
C Put it over on the table that's by the door.
D Put it over on the table that's by the door.

53

Suppose we have a trained first-order hidden Markov model for part-of-speech tagging with the parameters below (assume the value is 0 if the parameter is not shown in the tables).

Tag transition probabilities:

	DT	NN	VB	$\langle$/S$\rangle$
$\langle$S$\rangle$	0.5	0.3	0.2	0.0
DT	0.0	0.9	0.1	0.0
NN	0.1	0.2	0.3	0.4
VB	0.4	0.2	0.2	0.2

Word emission probabilities:

	the	can	see	$\langle$/s$\rangle$
DT	1.0	0.0	0.0	0.0
NN	0.0	0.9	0.1	0.0
VB	0.0	0.5	0.5	0.0
$\langle$/S$\rangle$	0.0	0.0	0.0	1.0

What is the predicted tag sequence for the sentence "the can can see $\langle$/s$\rangle$", decoded using greedy decoding, and what is the predicted probability of that tag sequence? Show your work. You can leave the answer as a product of decimals.

54

Suppose you want to build a system to perform part-of-speech tagging with a maximum entropy Markov model, and you have access to a working word sense disambiguation (WSD) system. How could you use the information provided by the WSD system to try to improve the performance of your model? Which parameter(s) of the model would you modify, and how? (2-3 sentences.)

55

Suppose we train a hidden Markov model to do part-of-speech tagging using the following training sequences:

- the/DT cat/NN is/VB in/PP the/DT box/NN
- the/DT cat/NN sat/VB on/PP the/DT mat/NN
- the/DT cat/NN is/VB black/JJ

We use a start tag $\langle S\rangle$, but no end tag, bigram tag transitions, and we smooth the transition probabilities using interpolation with $\lambda_1 = \lambda_2 = \frac{1}{2}$. What is the predicted tag sequence for "the black cat sat" under this model, and what is its probability? Show your work. You can leave your answer as a product of fractions.

52.

A These black cats detest those green peas.
B These black cats detest those green peas.
C Put it over on the table that's by the door.
D Put it over on the table that's by the door.

A. Constituent
B. Not a constituent
C. Not a constituent
D. Not a constituent

53. Hidden Markov Model POS Tagging

	DT	NN	VB	</S>			the	can	see	</s>
<S>	0.5	0.3	0.2	0.0		DT	1.0	0.0	0.0	0.0
DT	0.0	0.9	0.1	0.0		NN	0.0	0.9	0.1	0.0
NN	0.1	0.2	0.3	0.4		VB	0.0	0.5	0.5	0.0
VB	0.4	0.2	0.4	0.2		</S>	0.0	0.0	0.0	1.0

Predicted tag sequence for "the can can see </S>"
Greedy decoding

	the	can	can	see	</s>
DT	0.5*1				
NN	0				
VB	0				
</S>	0				

	the	can	can	see	</s>
DT	0.5	0			
NN	0	0.9*0.9*0.5 = 0.405			
VB	0	0.1*0.5*0.5 = 0.025			
</S>	0	0			

	the	can	can	see	</s>
DT	0.5	0	0		
NN	0	0.405	Max(0.405*0.2*0.9 = 0.0729 , 0.025*0.2*0.9 = 0.0045)		
VB	0	0.025	Max(0.405*0.3*0.5 = 0.06075, 0.025*0.2*0.5 = 0.0025)		
</S>	0	0	0		

	the	can	can	see	</s>
DT	0.5	0	0	0	
NN	0	0.405	0.0729	Max(0.0729*0.2*0.1 = 0.001458, 0.06075*0.2*0.1 = 0.00122)	
VB	0	0.025	0.06075	Max(0.0729*0.2*0.5 = 0.00729, 0.06075*0.2*0.5 = 0.00608)	
</S>	0	0	0	0	

	the	can	can	see	</s>
DT	0.5	0	0	0	0
NN	0	0.405	0.0729	0.001458	0
VB	0	0.025	0.06075	0.00729	0
</S>	0	0	0	0	Max(0.001458*0.4*1 = 0.0005832, 0.00729*0.2*1 = 0.00146)

	the	can	can	see	</s>
DT	0.5	0	0	0	0
NN	0	0.405	0.0729	0.001458	0
VB	0	0.025	0.06075	0.00729	0
</S>	0	0	0	0	0.00146
	DT	NN	NN	VB	</S>

4.

<S>	DT	NN	VB	PP	DT	NN	
<S>	DT	NN	VB	PP	DT	NN	
<S>	DT	NN	VB	JJ			

<u>TRANSITION PROBABILITIES</u>

	DT	NN	VB	PP	JJ
<S>	1	0	0	0	0
DT	0	1	0	0	0
NN	0	0	1	0	0
VB	0	0	0	0.67	0.33
PP	1	0	0	0	0
JJ	0	0	0	0	0

<u>UNIGRAM TRANSITION PROBABILITIES</u>

	c	p
<S>	3	3/19 = 0.1579
DT	5	5/19 = 0.2632
NN	5	5/19 = 0.2632
VB	3	3/19 = 0.1579
PP	2	2/19 = 0.1053
JJ	1	1/19 = 0.0526
	19	

<u>SMOOTHED TRANSITION PROBABILITIES</u>

	DT	NN	VB	PP	JJ
<S>	0.578947	0.078947	0.078947	0.078947	0.078947
DT	0.131579	0.631579	0.131579	0.131579	0.131579
NN	0.131579	0.131579	0.631579	0.131579	0.131579
VB	0.078947	0.078947	0.078947	0.413947	0.243947
PP	0.552632	0.052632	0.052632	0.052632	0.052632
JJ	0.026316	0.026316	0.026316	0.026316	0.026316

EMISSION PROBABILITIES

	the	cat	is	in	box	sat	on	mat	black
DT	1	0	0	0	0	0	0	0	0
NN	0	0.6	0	0	0.2	0	0	0.2	0
VB	0	0	0.67	0	0	0.33	0	0	0
PP	0	0	0	0.5	0	0	0.5	0	0
JJ	0	0	0	0	0	0	0	0	1

	the	black	cat	sat
DT	0.578			
NN	0			
VB	0			
PP	0			
JJ	0			

	the	black	cat	sat
DT	0.578	0		
NN	0	0		
VB	0	0		
PP	0	0		
JJ	0	0.578*0.1315 = 0.076		

	the	black	cat	sat
DT	0.578	0	0	
NN	0	0	0.076*0.0263*0.6 = 0.0012	
VB	0	0	0	
PP	0	0	0	
JJ	0	0.076	0	

	the	black	cat	sat
DT	0.578	0	0	0
NN	0	0	0.0012	0
VB	0	0	0	0.33*0.0012*0.6315 = 0.0003
PP	0	0	0	0
JJ	0	0.076	0	0
	DT	JJ	NN	VB

56

Suppose we train a language model on the following training sequences:

- the man saw the dog with the telescope
- the dog saw the man in the park
- the man with the telescope saw the park

We use linear interpolation between a bigram model and a unigram model, with $\lambda_1 = \lambda_2 = 0.5$. The bigram model is smoothed using Kneser-Ney smoothing with $\beta = 1$, and the unigram model is not smoothed. We do not use any start or end tokens. What is $p(\text{dog}|\text{the})$ under this model? Show your work. You can leave the answer as a sum of fractions.

57

Suppose we have two unigram probability distributions:

$$p(\text{cat}) = \tfrac{1}{2} \;\; p(\text{dog}) = \tfrac{1}{4}$$
$$p(\text{mat}) = \tfrac{1}{8} \;\; p(\text{log}) = \tfrac{1}{8}$$

and

$$q(\text{cat}) = \tfrac{1}{4} \;\; q(\text{dog}) = \tfrac{1}{8}$$
$$q(\text{mat}) = \tfrac{1}{8} \;\; q(\text{log}) = \tfrac{1}{4}$$

What is the cross-entropy $H(p, q)$? Show your work.

56.

Suppose we train a language model on the following training sequences:
the man saw the dog with the telescope
the dog saw the man in the park
the man with the telescope saw the park

p(dog|the)

Linear interpolation :
1. Bigram + Knesser-Ney smoothing (β=1)
2. Unigram Model

c(u,v)	the	man	saw	dog	with	telescope	in	park	SUM	c(u,v)>0
the	0	3	0	2	0	2	0	2	9	4
man	0	0	1	0	1	0	1	0	3	3
saw	3	0	0	0	0	0	0	0	3	1
dog	0	0	1	0	1	0	0	0	2	2
with	2	0	0	0	0	0	0	0	2	1
telescope	0	0	1	0	0	0	0	0	1	1
in	1	0	0	0	0	0	0	0	1	1
park	0	0	0	0	0	0	0	0	0	0
									21	13

<u>DISCOUNTED COUNTS</u>

										$\lambda(w_{i-1})$
c(u,v)	the	man	saw	dog	with	telescope	in	park		
the	0	2	0	1	0	1	0	1		0.44444
man	0	0	0	0	0	0	0	0		**1**
saw	2	0	0	0	0	0	0	0		0.33333
dog	0	0	0	0	0	0	0	0		1
with	1	0	0	0	0	0	0	0		0.5
telescope	0	0	0	0	0	0	0	0		0.5
in	0	0	0	0	0	0	0	0		1
park	0	0	0	0	0	0	0	0		0

P(dog|the)

λ(the) = 0.444

	c(u)	p
the	9	0.38
man	3	0.13
saw	3	0.13
dog	2	0.08
with	2	0.08
telescope	2	0.08
in	1	0.04
park	2	0.08
	24	1

P(dog|the)

= 0.5*[Ps(dog|the)] + 0.5*P(dog)
= 0.5*[c_discounted(the, dog)/c(the) + λ(man) * P_cont(dog)] + 0.5*(2/24)

c_discounted(the, dog) = 2-1 = 1
c(the) = 9

λ(the) = 4/9
P_cont(dog) = 2/13

= 0.5* [1/9 + (4/9)*(2/13)] + 0.5 * (2/24)

57)

$\sum q \neq 1$, q is not a probability distribution

$$H(p, q) = E_p[-\log(q)]$$

$$= -\left[\frac{1}{2}\log\left(\frac{1}{4}\right) + \frac{1}{4}\log\left(\frac{1}{8}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right) + \frac{1}{8}\log\left(\frac{1}{4}\right)\right]$$

$$= -\left[\frac{1}{2}(-2) + \frac{1}{4}(-3) + \frac{1}{8}(-3) + \frac{1}{8}(-2)\right]$$

$$= 1 + \frac{3}{4} + \frac{3}{8} + \frac{1}{4} = \frac{19}{8}$$

58

Suppose we have the following probabilistic context-free grammar:

S	$\rightarrow$	NP VP	1.0
VP	$\rightarrow$	VB NP	1.0
NP	$\rightarrow$	DT NN	0.5
NP	$\rightarrow$	DT NBAR	0.5
NBAR	$\rightarrow$	JJ NBAR	0.4
NBAR	$\rightarrow$	JJ NN	0.6
DT	$\rightarrow$	the	1.0
JJ	$\rightarrow$	old	0.8
JJ	$\rightarrow$	man	0.2
NN	$\rightarrow$	old	0.2
NN	$\rightarrow$	man	0.4
NN	$\rightarrow$	ships	0.4
VB	$\rightarrow$	man	0.5
VB	$\rightarrow$	ships	0.5

How many possible parses are there for the sentence “the old man the ships” under this grammar? Which has the highest probability, and what is that probability? Show your work.

59

Suppose we have a language model that uses discounting with Katz backoff to handle rare/out-of-vocabulary words. Is this strategy for handling rare/out-of-vocabulary words sufficient, or do we also need to use an unknown word token? Briefly explain your answer (1-2 sentences).

60

Which of the following underlined phrases is a constituent? Choose all that apply.

A These black cats detest those green peas.
B These black cats detest those green peas.
C Put it over on the table that’s by the door.
D Put it over on the table.

58.

	the	old	man	the	ships
the	DT 1.0				
old		JJ 0.8, NN 0.2			
man			JJ 0.2, NN 0.4, VB 0.5		
the				DT 1.0	
ships					NN 0.4, VB 0.5

	the	old	man	the	ships
the	DT 1.0	NP 0.5*1.0*0.2 = 0.1			
old		JJ 0.8, NN 0.2	NBAR 0.6* 0.8*0.4 = 0.192		
man			JJ 0.2, NN 0.4, VB 0.5	-	
the				DT 1.0	NP 0.5*1.0*0.4 = 0.2
ships					NN 0.4, VB 0.5

	the	old	man	the	ships
the	DT 1.0	NP 0.1	NP 0.5*1.0*0.192 = 0.096		
old		JJ 0.8, NN 0.2	NBAR 0.192	-	
man			JJ 0.2, NN 0.4, VB 0.5	-	VP 1.0*0.5*0.2 = 0.1
the				DT 1.0	NP 0.2
ships					NN 0.4, VB 0.5

	the	old	man	the	ships
the	DT 1.0	NP 0.1	NP 0.096	-	
old		JJ 0.8, NN 0.2	NBAR 0.192	-	-
man			JJ 0.2, NN 0.4, VB 0.5	-	VP 0.1
the				DT 1.0	NP 0.2
ships					NN 0.4, VB 0.5

	the	old	man	the	ships
the	DT 1.0	NP 0.1	NP 0.096	-	S 1.0*0.1*0.1 = 0.01
old		JJ 0.8, NN 0.2	NBAR 0.192	-	-
man			JJ 0.2, NN 0.4, VB 0.5	-	VP 0.1
the				DT 1.0	NP 0.2
ships					NN 0.4, VB 0.5

There is 1 parse for the sentence "the old man the ships".

Probability for the parse : 0.01

60.

A These black cats detest those green peas.
B These black cats detest those green peas.
C Put it over on the table that's by the door.
D Put it over on the table.

A. Constituent
B. Constituent
C. Not a constituent
D. Constituent

61

Suppose we train a Naive Bayes classifier on the following training sequences (the letter after the comma is the class label):

- the cat sat on the mat, A
- the cat sat in the hat, A
- the dog sat on the log, B
- the dog sat on the cat, B
- the fish sat in the dish, C
- the fish in the hat sat, C

Our classifier uses bigram count features, no start or end tokens, and Laplace smoothing with $\delta = 1$. Given the test sentence "the cat in the hat," what is the predicted probability of this sentence belonging to class C under this model? Show your work. You can leave the answer as a product of fractions.

62

Suppose we have a training corpus of two sentences:

1.	souris	blanche	2.	souris
	white	mouse		mouse

Additionally, suppose we already have a trained IBM Model 1 statistical machine translation system, and now we want to train an IBM Model 2. The parameters $q(j|i, l, m)$ are initialized uniformly, and the parameters $t(f_i|e_j)$ are initialized using the Model 1 as follows:

- $t(\text{souris}|\text{white}) = \frac{1}{2}$
- $t(\text{souris}|\text{mouse}) = \frac{3}{4}$
- $t(\text{blanche}|\text{white}) = \frac{1}{2}$
- $t(\text{blanche}|\text{mouse}) = \frac{1}{4}$

Perform one iteration of expectation maximization to update the parameters using the training corpus. What is $t(\text{souris}|\text{white})$ at the end of this iteration? Show your work.

63

Suppose we have a language model that uses Laplace smoothing to handle rare/out-of-vocabulary words. Is this strategy for handling rare/out-of-vocabulary words sufficient, or do we also need to use an unknown word token? Briefly explain your answer (1-2 sentences).

61. Bigram count features
Laplace smoothing

A : COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	2	0	0	1	0	1	0	0	0	0	4
cat	0	0	2	0	0	0	0	0	0	0	0	2
sat	0	0	0	1	0	1	0	0	0	0	0	2
on	1	0	0	0	0	0	0	0	0	0	0	1
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	1	0	0	0	0	0	0	0	0	0	0	1
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	2	2	1	1	1	1	0	0	0	0	10

SMOOTHED A COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	3	1	1	2	1	2	1	1	1	1	15
cat	1	1	3	1	1	1	1	1	1	1	1	13
sat	1	1	1	2	1	2	1	1	1	1	1	13
on	2	1	1	1	1	1	1	1	1	1	1	12
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	2	1	1	1	1	1	1	1	1	1	1	12
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	13	13	12	12	12	12	11	11	11	11	131

A : PROBABILITY MATRIX

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.00763	0.0229	0.0076	0.0076	0.0153	0.0076	0.0153	0.007633	0.0076	0.007633588	0.0076	0.1145
cat	0.00763	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0992
sat	0.00763	0.0076	0.0076	0.0153	0.0076	0.0153	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0992
on	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0916
mat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
in	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0916
hat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
dog	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
log	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
fish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
dish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
	0.09924	0.0992	0.0992	0.0916	0.0916	0.0916	0.0916	0.083969	0.084	0.083969466	0.084	1

<table>
<tr><td>the</td><td>cat</td><td>in</td><td>the</td><td>hat</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>0.0229</td><td>0.0076</td><td>0.0153</td><td>0.0153</td><td>4.0747E-08</td><td>0.3333</td><td>1.35823 E-08</td><td></td><td>SUM</td><td>2.26372 E-08</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td>P(A|d)</td><td>0.6</td><td></td><td></td><td></td></tr>
</table>

B : COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	1	0	0	0	0	0	2	1	0	0	4
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	0	0	0	2	0	0	0	0	0	0	0	2
on	2	0	0	0	0	0	0	0	0	0	0	2
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	0	0	0	0	0	0	0	0	0	0	0	0
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	2	0	0	0	0	0	0	0	0	2
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	2	2	0	0	0	2	1	0	0	10

B : SMOOTHED COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	2	1	1	1	1	1	3	2	1	1	15
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	1	1	1	3	1	1	1	1	1	1	1	13
on	3	1	1	1	1	1	1	1	1	1	1	13
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	1	1	1	1	1	1	1	1	1	1	1	11
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	3	1	1	1	1	1	1	1	1	13
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	12	13	13	11	11	11	13	12	11	11	131

B : PROBABILITY MATRIX

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0229	0.0153	0.0076	0.0076	0.1145
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0076	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
on	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
hat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dog	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.0992	0.0916	0.0992	0.0992	0.084	0.084	0.084	0.0992	0.0916	0.084	0.084	1

the	cat	in	the	hat			
	0.0153	0.0076	0.0076	0.0076	7E-09	0.3333	2E-09
						P(B|d)	**0.1**

C : COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	0	0	0	0	0	1	0	0	2	1	4
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	0	0	0	0	0	1	0	0	0	0	0	1
on	0	0	0	0	0	0	0	0	0	0	0	0
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	2	0	0	0	0	0	0	0	0	0	0	2
hat	0	0	1	0	0	0	0	0	0	0	0	1
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	1	0	0	1	0	0	0	0	0	2
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	2	0	0	2	1	0	0	2	1	10

C : SMOOTHED COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	1	1	1	1	1	2	1	1	3	2	15
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	1	1	1	1	1	2	1	1	1	1	1	12
on	1	1	1	1	1	1	1	1	1	1	1	11
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	3	1	1	1	1	1	1	1	1	1	1	13
hat	1	1	2	1	1	1	1	1	1	1	1	12
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	2	1	1	2	1	1	1	1	1	13
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	11	13	11	11	13	12	11	11	13	12	131

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0229	0.0153	0.1145
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
on	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
hat	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
dog	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0153	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.0992	0.084	0.0992	0.084	0.084	0.0992	0.0916	0.084	0.084	0.0992	0.0916	1

the	cat	in	the	hat			
	0.0076	0.0076	0.0229	0.0153	2E-08	0.3333	7E-09
						P(C|d)	**0.3**

62.

EM to update parameters and find t(souris|white)

INITIAL PARAMETERS:

t	souris	blanche
white	1/2	1/2
mouse	3/4	1/4

We have two sentences :

For k = 1 : (m = 2, l=2)
For k = 2 : (m = 1, l=1)

Initial q's are initialized uniformly

q(j|i,l,m)
j - ENGLISH WORD
i - FRENCH WORD

q(1|1,1,1)	1		
q(1|1,2,2)	1/2	q(2|1,2,2)	1/2
q(1|2,2,2)	1/2	q(2|2,2,2)	1/2

NEXT STEP:

δ(k,i,j)

δ(2,1,1)	1	
δ(1,1,1)	q(1|1,2,2) t(souris|white) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] = (1/2)(1/2)/ [(1/2)(1/2) + (1/2)(3/4)]	2/5
δ(1,1,2)	q(2|1,2,2) t(souris|mouse) / [(q(1|1,2,2) t(souris|white) + q(2|1,2,2) t(souris|mouse))] =(1/2)(3/4)/ [(1/2)(1/2) + (1/2)(3/4)]	3/5
δ(1,2,1)	q(1|2,2,2) t(blanche|white) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/2) / [(1/2)(1/2) + (1/2)(1/4)]	2/3
δ(1,2,2)	q(2|2,2,2) t(blanche|mouse) / [q(1|2,2,2) t(blanche|white) + q(2|2,2,2) t(blanche|mouse))] = (1/2)(1/4) / [(1/2)(1/2) + (1/2)(1/4)]	1/3

c(white, souris)	2/5		c(white)	2/5 + 2/3		t(souris|white)	c(white,souris)/c(white)
c(white, blanche)	2/3		c(mouse)	3/5 + 1/3 + 1			(2/5)/ (2/5) + (2/3)
c(mouse, souris)	3/5 + 1						(6/16)
c(mouse, blanche)	1/3						

64

Suppose you want to build a system to perform part-of-speech tagging with a first-order hidden Markov model, and you have access to a working word sense disambiguation (WSD) system. How could you use the information provided by the WSD system to try to improve the performance of your model? Which parameter(s) of the model would you modify, and how? (2-3 sentences.)

65

Suppose we have two unigram probability distributions:

$$p(\text{cat}) = \tfrac{1}{2} \quad p(\text{dog}) = \tfrac{1}{4}$$
$$p(\text{mat}) = \tfrac{1}{8} \quad p(\text{log}) = \tfrac{1}{8}$$

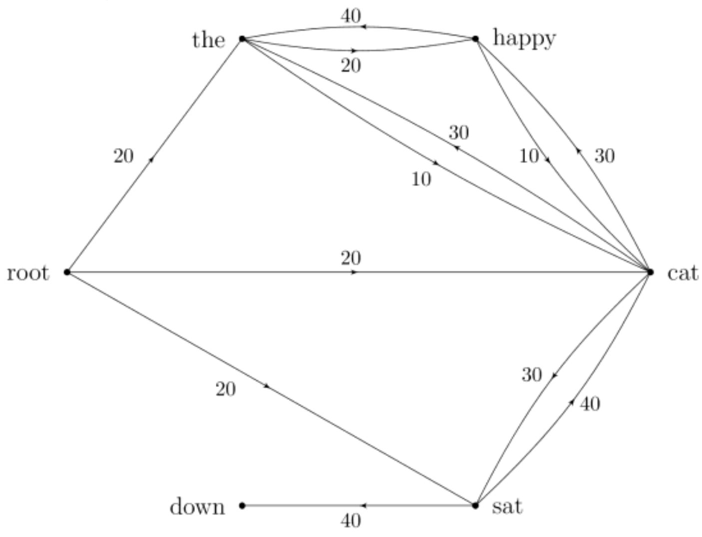
and

$$q(\text{cat}) = \tfrac{1}{4} \quad q(\text{dog}) = \tfrac{1}{2}$$
$$q(\text{mat}) = \tfrac{1}{8} \quad q(\text{log}) = \tfrac{1}{8}$$

What is the cross-entropy $H(p, q)$? Show your work.

66

Suppose we have the following graph model of candidate dependencies for the sentence "the happy cat sat down":



You can assume that any edges not shown have score 0. What is the highest-scoring dependency parse tree based on this graph (you can give your answer using arrows (A $\rightarrow$ B) or ordered pairs (A, B)), and what is its total score? Show your work.

65)

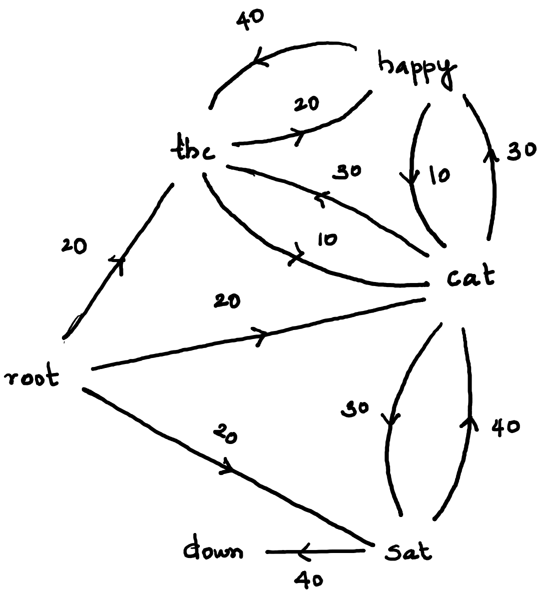
$$H(p,q) = E_p\left[-\log(q)\right]$$

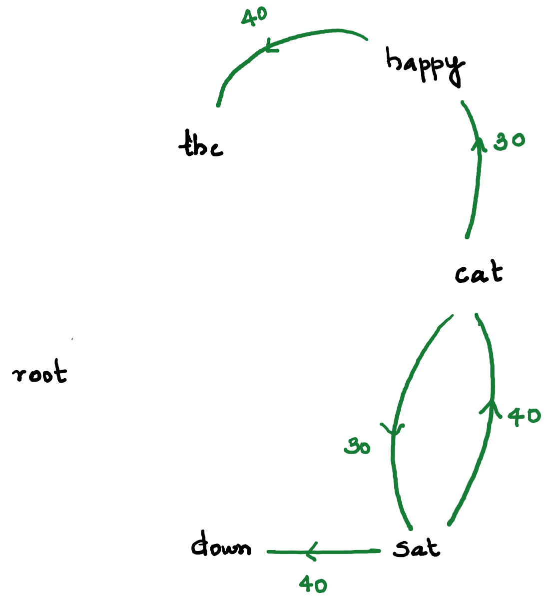
$$= -\left[\frac{1}{2}\log\left(\frac{1}{4}\right) + \frac{1}{4}\log\left(\frac{1}{2}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right)\right]$$

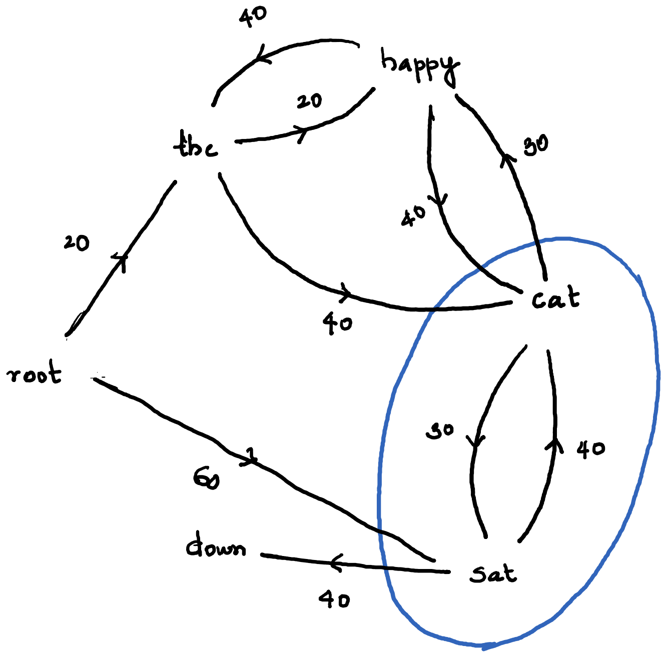
$$= -\left[\frac{1}{2}(-2) + \frac{1}{4}(-1) + \frac{1}{4}(-3)\right]$$

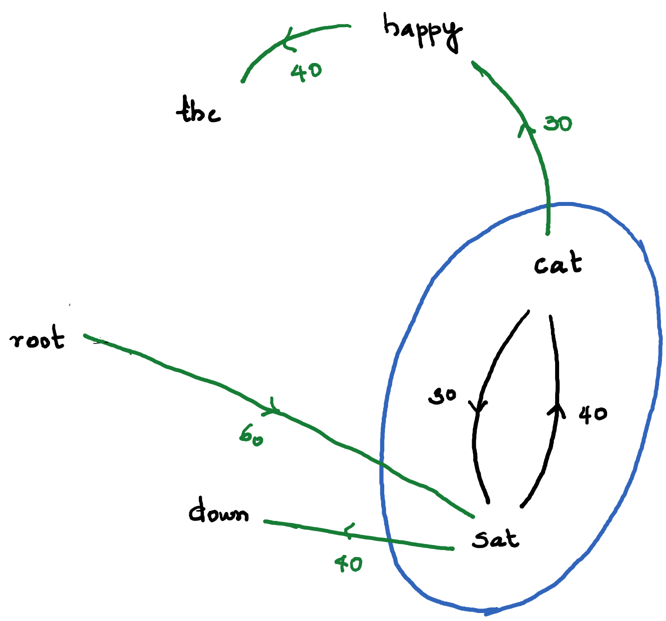
$$= 1 + \frac{1}{4} + \frac{3}{4} = 2$$

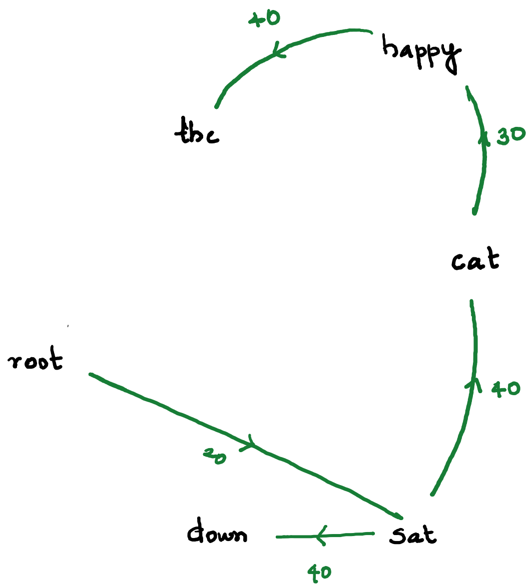
66)


40
happy
20
the
30
10
30
20
cat
20
root
20
30
40
down
sat
40


40
happy
the
30
cat
root
40
30
down
sat
40


40
happy
20
the
30
40
20
cat
40
root
30
40
60
down
sat
40


happy
40
the
30
cat
root
30
40
60
down
sat
40

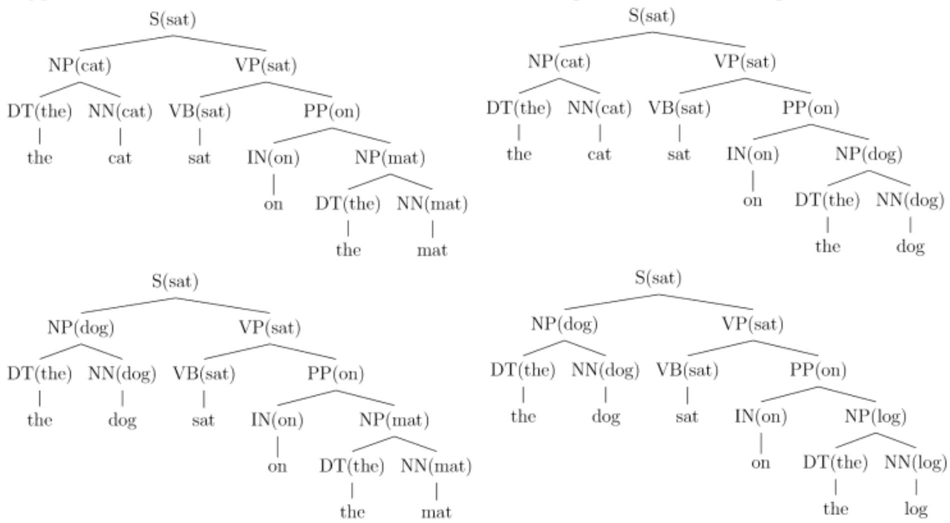


Total Score : 40 + 30 + 40 + 40 + 20

= 170

67

Suppose we want to estimate a lexicalized PCFG using these four training trees:



What is the probability of the rule S(sat) $\rightarrow$ NP(log) VP(sat), estimated using Charniak's method with uniform λs for smoothing? You can assume that the position of the head is fixed for each rule, so there is only one rule, not two different rules depending on head position. Show your work.

68

Suppose we train a Naive Bayes classifier on the following training sequences (the letter after the comma is the class label):

- the cat sat on the mat, A
- the cat sat in the hat, A
- the dog sat on the log, B
- the dog sat on the cat, B
- the fish sat in the dish, C
- the fish in the hat sat, C

Our classifier uses bigram count features, no start or end tokens, and Laplace smoothing with $\delta = 1$. Given the test sentence "the cat in the hat," what is the predicted probability of this sentence belonging to class B under this model? Show your work. You can leave the answer as a product of fractions.

67)

$$p\left(S(sat) \rightarrow NP(log)\ VP(sat)\right)$$

$$= \underbrace{p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right)}_{\text{I}}\ \underbrace{p\left(log \mid S(sat) \rightarrow NP\ VP(sat)\right)}_{\text{II}}$$

Considering I : .

$$p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) =$$

$$\lambda_1 P_{MLE}\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) + \lambda_2 P_{MLE}\left(S \rightarrow NP\,VP \mid S\right)$$

We have $\lambda_1 = \lambda_2 = 1/2$

$$P_{MLE}\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) = 1$$

$$P_{MLE}\left(S \rightarrow NP\,VP \mid S\right) = 1$$

For I we have $p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) = 1$

Considering II : $p\left(log \mid S(sat) \rightarrow NP\ VP(sat)\right)$

$$= \lambda_1 P_{MLE}\left(log \mid S(sat) \rightarrow NP\ VP(sat)\right) + \lambda_2 P_{MLE}\left(log \mid S \rightarrow NP\,VP\right) + \lambda_3 P_{MLE}\left(log \mid NP\right)$$

We have $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$ & $P_{MLE}\left(log \mid S(sat) \rightarrow NP\,VP(sat)\right) = 0$

$$P_{MLE}\left(log \mid S \rightarrow NP\,VP\right) = 0, \quad P_{MLE}\left(log \mid NP\right) = \frac{1}{8}$$

$$\Rightarrow p\left(S(sat) \rightarrow NP(log)\ VP(sat)\right) = 1\left(\frac{1}{3} \times \frac{1}{8}\right) = \boxed{\frac{1}{24}}$$

61. Bigram count features
Laplace smoothing

A : COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	2	0	0	1	0	1	0	0	0	0	4
cat	0	0	2	0	0	0	0	0	0	0	0	2
sat	0	0	0	1	0	1	0	0	0	0	0	2
on	1	0	0	0	0	0	0	0	0	0	0	1
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	1	0	0	0	0	0	0	0	0	0	0	1
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	2	2	1	1	1	1	0	0	0	0	10

SMOOTHED A COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	3	1	1	2	1	2	1	1	1	1	15
cat	1	1	3	1	1	1	1	1	1	1	1	13
sat	1	1	1	2	1	2	1	1	1	1	1	13
on	2	1	1	1	1	1	1	1	1	1	1	12
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	2	1	1	1	1	1	1	1	1	1	1	12
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	13	13	12	12	12	12	11	11	11	11	131

A : PROBABILITY MATRIX

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.00763	0.0229	0.0076	0.0076	0.0153	0.0076	0.0153	0.007633	0.0076	0.007633588	0.0076	0.1145
cat	0.00763	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0992
sat	0.00763	0.0076	0.0076	0.0153	0.0076	0.0153	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0992
on	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0916
mat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
in	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0916
hat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
dog	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
log	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
fish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
dish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
	0.09924	0.0992	0.0992	0.0916	0.0916	0.0916	0.0916	0.083969	0.084	0.083969466	0.084	1

the	cat	in	the	hat						
	0.0229	0.0076	0.0153	0.0153	4.0747E-08	0.3333	1.35823E-08		SUM	2.26372E-08
						P(A|d)	**0.6**			

B : COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	1	0	0	0	0	0	2	1	0	0	4
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	0	0	0	2	0	0	0	0	0	0	0	2
on	2	0	0	0	0	0	0	0	0	0	0	2
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	0	0	0	0	0	0	0	0	0	0	0	0
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	2	0	0	0	0	0	0	0	0	2
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	2	2	0	0	0	2	1	0	0	10

B : SMOOTHED COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	2	1	1	1	1	1	3	2	1	1	15
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	1	1	1	3	1	1	1	1	1	1	1	13
on	3	1	1	1	1	1	1	1	1	1	1	13
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	1	1	1	1	1	1	1	1	1	1	1	11
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	3	1	1	1	1	1	1	1	1	13
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	12	13	13	11	11	11	13	12	11	11	131

B : PROBABILITY MATRIX

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0229	0.0153	0.0076	0.0076	0.1145
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0076	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
on	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
hat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dog	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.0992	0.0916	0.0992	0.0992	0.084	0.084	0.084	0.0992	0.0916	0.084	0.084	1

the	cat	in	the	hat			
	0.0153	0.0076	0.0076	0.0076	7E-09	0.3333	2E-09
						P(B|d)	**0.1**

C : COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	0	0	0	0	0	1	0	0	2	1	4
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	0	0	0	0	0	1	0	0	0	0	0	1
on	0	0	0	0	0	0	0	0	0	0	0	0
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	2	0	0	0	0	0	0	0	0	0	0	2
hat	0	0	1	0	0	0	0	0	0	0	0	1
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	1	0	0	1	0	0	0	0	0	2
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	2	0	0	2	1	0	0	2	1	10

C : SMOOTHED COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	1	1	1	1	1	2	1	1	3	2	15
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	1	1	1	1	1	2	1	1	1	1	1	12
on	1	1	1	1	1	1	1	1	1	1	1	11
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	3	1	1	1	1	1	1	1	1	1	1	13
hat	1	1	2	1	1	1	1	1	1	1	1	12
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	2	1	1	2	1	1	1	1	1	13
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	11	13	11	11	13	12	11	11	13	12	131

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0229	0.0153	0.1145
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
on	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
hat	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
dog	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0153	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.0992	0.084	0.0992	0.084	0.084	0.0992	0.0916	0.084	0.084	0.0992	0.0916	1

the	cat	in	the	hat			
	0.0076	0.0076	0.0229	0.0153	2E-08	0.3333	7E-09
						P(C|d)	**0.3**

69

Suppose we have a language model that has an unknown word token `unk` to handle rare/out-of-vocabulary words. Is this strategy for handling rare/out-of-vocabulary words sufficient, or do we also need to use smoothing? Briefly explain your answer (1-2 sentences).

70

Suppose we have two unigram probability distributions:

$$p(\text{cat}) = \tfrac{1}{2} \;\; p(\text{dog}) = \tfrac{1}{4}$$
$$p(\text{mat}) = \tfrac{1}{8} \;\; p(\text{log}) = \tfrac{1}{8}$$

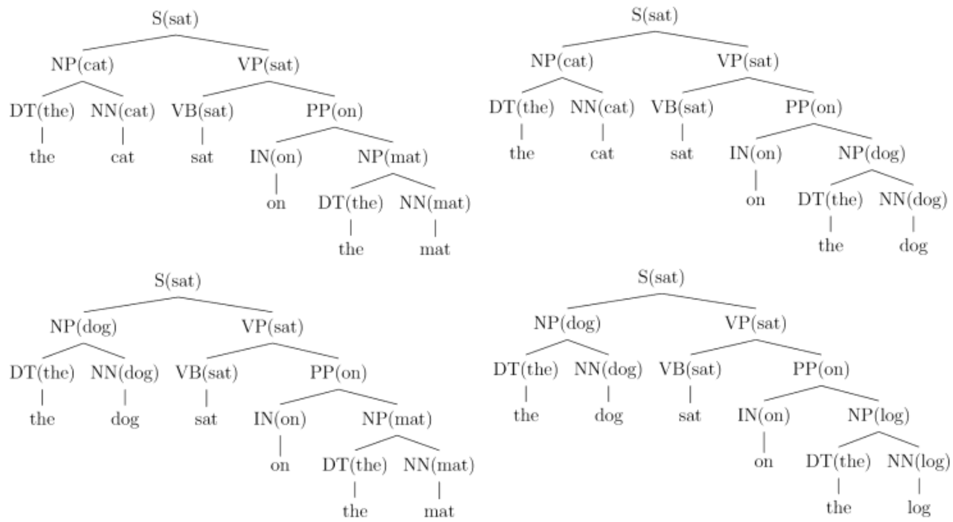
and

$$q(\text{cat}) = \tfrac{1}{2} \;\; q(\text{dog}) = \tfrac{1}{8}$$
$$q(\text{mat}) = \tfrac{1}{4} \;\; q(\text{log}) = \tfrac{1}{8}$$

What is the cross-entropy $H(p, q)$? Show your work.

71

Suppose we want to estimate a lexicalized PCFG using these four training trees:



What is the probability of the rule S(sat) $\rightarrow$ NP(dog) VP(sat), estimated using Charniak's method with uniform λs for smoothing? You can assume that the position of the head is fixed for each rule, so there is only one rule, not two different rules depending on head position. Show your work.

70)

$$H(p,q) = E_p\left[-\log(q)\right]$$

$$= -\left[\frac{1}{2}\log\left(\frac{1}{2}\right) + \frac{1}{4}\log\left(\frac{1}{8}\right) + \frac{1}{8}\log\left(\frac{1}{4}\right) + \frac{1}{8}\log\left(\frac{1}{8}\right)\right]$$

$$= -\left[\frac{1}{2}(-1) + \frac{1}{4}(-3) + \frac{1}{8}(-2) + \frac{1}{8}(-3)\right]$$

$$= -\left[-\frac{1}{2} - \frac{3}{4} - \frac{1}{4} - \frac{3}{8}\right]$$

$$= \frac{3}{2} + \frac{3}{8} = \frac{12}{8} = \frac{3}{2}$$

71)

$$p\left(S(sat) \rightarrow NP(dog)\ VP(sat)\right)$$

$$= \underbrace{p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right)}_{\text{I}}\ \underbrace{p\left(dog \mid S(sat) \rightarrow NP\ VP(sat)\right)}_{\text{II}}$$

Considering I :

$$p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) =$$

$$\lambda_1 P_{MLE}\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) + \lambda_2 P_{MLE}\left(S \rightarrow NP\ VP \mid S\right)$$

We have $\lambda_1 = \lambda_2 = 1/2$

$$P_{MLE}\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) = 1$$

$$P_{MLE}\left(S \rightarrow NP\ VP \mid S\right) = 1$$

For I we have $p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat)\right) = 1$

Considering II :

$$p\left(dog \mid S(sat) \rightarrow NP\ VP(sat)\right) =$$

$$\lambda_1 P_{MLE}\left(dog \mid S(sat) \rightarrow NP\ VP(sat)\right) + \lambda_2 P_{MLE}\left(dog \mid S \rightarrow NP\ VP\right) + \lambda_3 P_{MLE}(dog \mid NP)$$

We have $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$

$$p\left(\text{dog} \mid S(\text{sat}) \rightarrow NP\ VP(\text{sat})\right) = \frac{1}{2}$$

$$p\left(\text{dog} \mid S \rightarrow NP\ VP\right) = \frac{1}{2}$$

$$p\left(\text{dog} \mid NP\right) = \frac{3}{8}$$

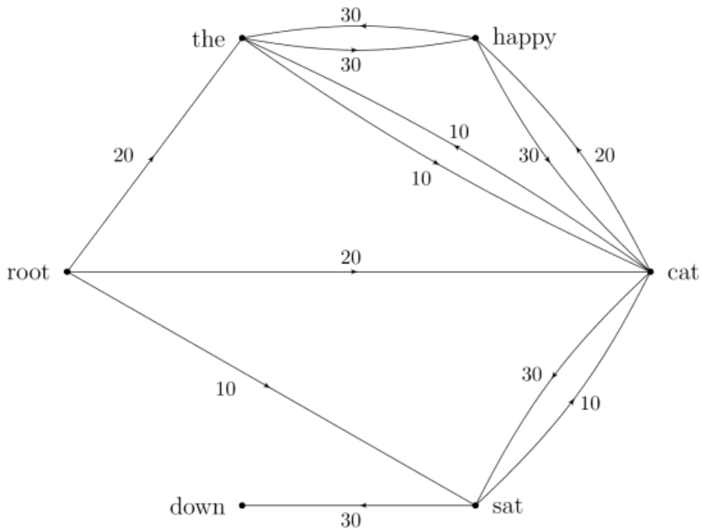
For II : $\frac{1}{3}\left[\frac{1}{2} + \frac{1}{2} + \frac{3}{8}\right]$

$$= \frac{1}{3} + \frac{1}{8} = \frac{11}{24}$$

$$p\left(S(\text{sat}) \longrightarrow NP(\text{dog})\ VP(\text{sat})\right) = \boxed{\frac{11}{24}}$$

72

Suppose we have the following graph model of candidate dependencies for the sentence "the happy cat sat down":



You can assume that any edges not shown have score 0. What is the highest-scoring dependency parse tree based on this graph (you can give your answer using arrows (A → B) or ordered pairs (A, B)), and what is its total score? Show your work.

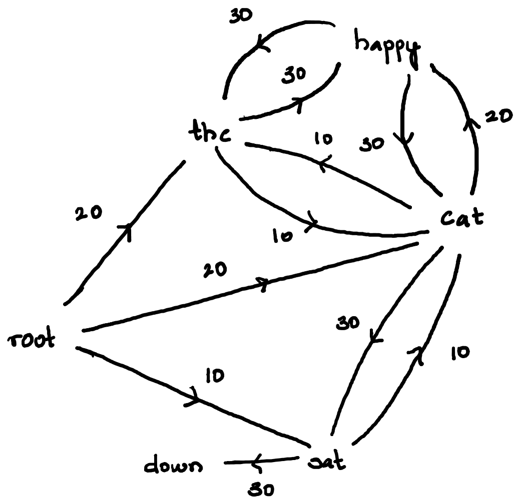
73

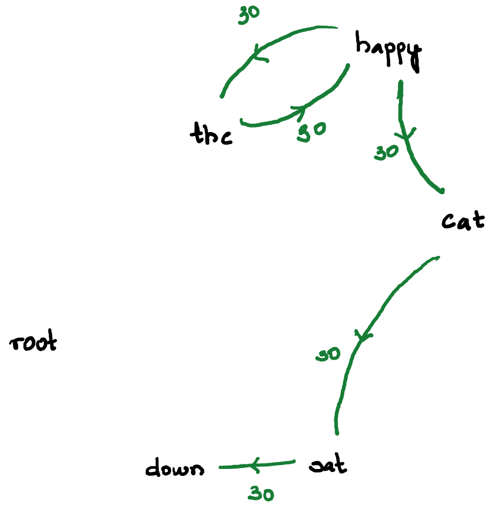
Suppose we train a Naive Bayes classifier on the following training sequences (the letter after the comma is the class label):

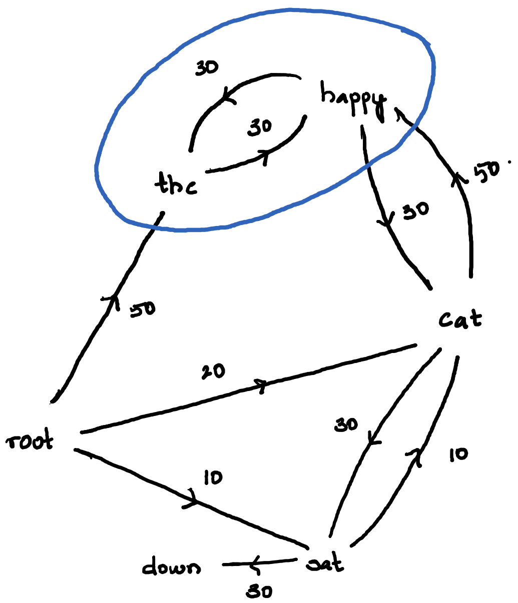
- the cat sat on the mat, A
- the cat sat in the hat, A
- the dog sat on the log, B
- the dog sat on the cat, B
- the fish sat in the dish, C
- the fish in the hat sat, C

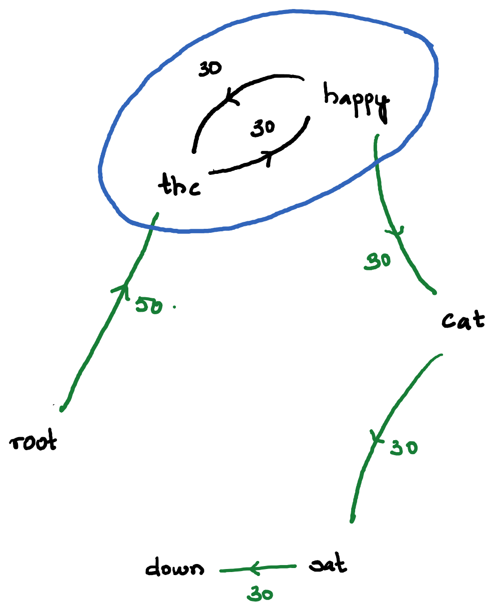
Our classifier uses bigram count features, no start or end tokens, and Laplace smoothing with $\delta = 1$. Given the test sentence "the cat in the hat," what is the predicted probability of this sentence belonging to class A under this model? Show your work. You can leave the answer as a product of fractions.

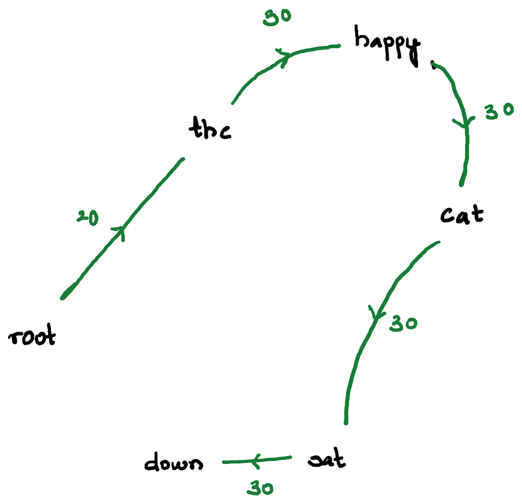
72)


30
happy
30
the
10
30
20
20
cat
10
20
root
30
10
10
down
sat
30


30
happy
the
30
30
cat
root
30
down
sat
30


30
happy
30
the
30
50
cat
50
20
root
30
10
10
down
sat
30


30
happy
30
the
30
50
cat
root
30
down
sat
30



Total Score : $20 + 30 + 30 + 30 + 30$

$= 140$

61. Bigram count features
    Laplace smoothing

A : COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	2	0	0	1	0	1	0	0	0	0	4
cat	0	0	2	0	0	0	0	0	0	0	0	2
sat	0	0	0	1	0	1	0	0	0	0	0	2
on	1	0	0	0	0	0	0	0	0	0	0	1
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	1	0	0	0	0	0	0	0	0	0	0	1
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	2	2	1	1	1	1	0	0	0	0	10

SMOOTHED A COUNTS

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	3	1	1	2	1	2	1	1	1	1	15
cat	1	1	3	1	1	1	1	1	1	1	1	13
sat	1	1	1	2	1	2	1	1	1	1	1	13
on	2	1	1	1	1	1	1	1	1	1	1	12
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	2	1	1	1	1	1	1	1	1	1	1	12
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	13	13	12	12	12	12	11	11	11	11	131

A : PROBABILITY MATRIX

A	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.00763	0.0229	0.0076	0.0076	0.0153	0.0076	0.0153	0.007633	0.0076	0.007633588	0.0076	0.1145
cat	0.00763	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0992
sat	0.00763	0.0076	0.0076	0.0153	0.0076	0.0153	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0992
on	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0916
mat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
in	0.01527	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.0916
hat	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
dog	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
log	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
fish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
dish	0.00763	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.007633	0.0076	0.007633588	0.0076	0.084
	0.09924	0.0992	0.0992	0.0916	0.0916	0.0916	0.0916	0.083969	0.084	0.083969466	0.084	1

the	cat	in	the	hat						
	0.0229	0.0076	0.0153	0.0153	4.0747E-08	0.3333	1.35823E-08		SUM	2.26372E-08
						P(A|d)	**0.6**			

B : COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	1	0	0	0	0	0	2	1	0	0	4
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	0	0	0	2	0	0	0	0	0	0	0	2
on	2	0	0	0	0	0	0	0	0	0	0	2
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	0	0	0	0	0	0	0	0	0	0	0	0
hat	0	0	0	0	0	0	0	0	0	0	0	0
dog	0	0	2	0	0	0	0	0	0	0	0	2
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	0	0	0	0	0	0	0	0	0	0
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	2	2	0	0	0	2	1	0	0	10

B : SMOOTHED COUNTS

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	2	1	1	1	1	1	3	2	1	1	15
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	1	1	1	3	1	1	1	1	1	1	1	13
on	3	1	1	1	1	1	1	1	1	1	1	13
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	1	1	1	1	1	1	1	1	1	1	1	11
hat	1	1	1	1	1	1	1	1	1	1	1	11
dog	1	1	3	1	1	1	1	1	1	1	1	13
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	1	1	1	1	1	1	1	1	1	11
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	12	13	13	11	11	11	13	12	11	11	131

B : PROBABILITY MATRIX

B	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0229	0.0153	0.0076	0.0076	0.1145
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0076	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
on	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
hat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dog	0.0076	0.0076	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.0992	0.0916	0.0992	0.0992	0.084	0.084	0.084	0.0992	0.0916	0.084	0.084	1

the	cat	in	the	hat			
	0.0153	0.0076	0.0076	0.0076	7E-09	0.3333	2E-09
						P(B|d)	**0.1**

C : COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0	0	0	0	0	0	1	0	0	2	1	4
cat	0	0	0	0	0	0	0	0	0	0	0	0
sat	0	0	0	0	0	1	0	0	0	0	0	1
on	0	0	0	0	0	0	0	0	0	0	0	0
mat	0	0	0	0	0	0	0	0	0	0	0	0
in	2	0	0	0	0	0	0	0	0	0	0	2
hat	0	0	1	0	0	0	0	0	0	0	0	1
dog	0	0	0	0	0	0	0	0	0	0	0	0
log	0	0	0	0	0	0	0	0	0	0	0	0
fish	0	0	1	0	0	1	0	0	0	0	0	2
dish	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	2	0	0	2	1	0	0	2	1	10

C : SMOOTHED COUNTS

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	1	1	1	1	1	1	2	1	1	3	2	15
cat	1	1	1	1	1	1	1	1	1	1	1	11
sat	1	1	1	1	1	2	1	1	1	1	1	12
on	1	1	1	1	1	1	1	1	1	1	1	11
mat	1	1	1	1	1	1	1	1	1	1	1	11
in	3	1	1	1	1	1	1	1	1	1	1	13
hat	1	1	2	1	1	1	1	1	1	1	1	12
dog	1	1	1	1	1	1	1	1	1	1	1	11
log	1	1	1	1	1	1	1	1	1	1	1	11
fish	1	1	2	1	1	2	1	1	1	1	1	13
dish	1	1	1	1	1	1	1	1	1	1	1	11
	13	11	13	11	11	13	12	11	11	13	12	131

C	the	cat	sat	on	mat	in	hat	dog	log	fish	dish	SUM
the	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0229	0.0153	0.1145
cat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
sat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
on	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
mat	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
in	0.0229	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
hat	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0916
dog	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
log	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
fish	0.0076	0.0076	0.0153	0.0076	0.0076	0.0153	0.0076	0.0076	0.0076	0.0076	0.0076	0.0992
dish	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.0076	0.084
	0.0992	0.084	0.0992	0.084	0.084	0.0992	0.0916	0.084	0.084	0.0992	0.0916	1

the	cat	in	the	hat			
	0.0076	0.0076	0.0229	0.0153	2E-08	0.3333	7E-09
						P(C|d)	**0.3**

04 December 2020 17:07

74

Suppose we want to estimate a lexicalized PCFG using these four training trees:

[S(sat) [NP(cat) [DT(the) the] [NN(cat) cat]] [VP(sat) [VB(sat) sat] [PP(on) [IN(on) on] [NP(mat) [DT(the) the] [NN(mat) mat]]]]]

[S(sat) [NP(cat) [DT(the) the] [NN(cat) cat]] [VP(sat) [VB(sat) sat] [PP(on) [IN(on) on] [NP(dog) [DT(the) the] [NN(dog) dog]]]]]

[S(sat) [NP(dog) [DT(the) the] [NN(dog) dog]] [VP(sat) [VB(sat) sat] [PP(on) [IN(on) on] [NP(mat) [DT(the) the] [NN(mat) mat]]]]]

[S(sat) [NP(dog) [DT(the) the] [NN(dog) dog]] [VP(sat) [VB(sat) sat] [PP(on) [IN(on) on] [NP(log) [DT(the) the] [NN(log) log]]]]]

What is the probability of the rule S(sat) $\rightarrow$ NP(cat) VP(sat), estimated using Charniak's method with uniform λs for smoothing? You can assume that the position of the head is fixed for each rule, so there is only one rule, not two different rules depending on head position. Show your work.

74)

Probability of the rule : $P\left(S(sat) \rightarrow NP(cat)\ VP(sat) \right)$

$$= \underbrace{P\left(S(sat) \rightarrow NP\ \ VP(sat) \mid S(sat) \right)}_{\text{I}} \quad \underbrace{P\left(cat \mid S(sat) \rightarrow NP\ VP(sat) \right)}_{\text{II}}$$

Considering I :

$$p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat) \right) =$$

$$\lambda_1 P_{MLE}\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat) \right) + \lambda_2 P_{MLE}\left(S \rightarrow NP\ VP \mid S \right)$$

We have $\lambda_1 = \lambda_2 = 1/2$

$$P_{MLE}\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat) \right) = 1$$

$$P_{MLE}\left(S \rightarrow NP\ VP \mid S \right) = 1$$

For I $\quad p\left(S(sat) \rightarrow NP\ VP(sat) \mid S(sat) \right) = 1/2 + 1/2 = \boxed{1}$

Considering II :

$$P\left(cat \mid S(sat) \rightarrow NP\ VP(sat) \right) =$$

$$\lambda_1 P_{MLE}\left(cat \mid S(sat) \rightarrow NP\ VP(sat) \right) + \lambda_2 P_{MLE}\left(cat \mid S \rightarrow NP\ VP \right) + \lambda_3 P_{MLE}\left(cat \mid NP \right)$$

We have $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$

$$P_{MLE}\left(\text{cat} \mid S(\text{sat}) \rightarrow NP\ VP(\text{sat})\right) = \frac{1}{2}$$

$$P_{MLE}\left(\text{cat} \mid S \rightarrow NP\ VP\right) = \frac{1}{2}$$

$$P_{MLE}\left(\text{cat} \mid NP\right) = \frac{1+1}{2+2+2+2} = \frac{1}{4}$$

For II we have

$$\frac{1}{3}\left(\frac{1}{2} + \frac{1}{2} + \frac{1}{4}\right) = \frac{5}{12}$$

So,

$$p\left(s(\text{sat}) \longrightarrow NP(\text{cat})\ VP(\text{sat})\right) = 1\left(\frac{5}{12}\right) = \frac{5}{12}$$