

CS6320, Fall 2016
Dr. Mithun Balakrishna
Homework 2
Due October 9th, 2016 11:59pm

A. Submission Instructions:

- Submit your solutions via eLearning.
- Please submit a single zip file with the following files:
 - For programming questions:
 - Source code file(s) in C/C++, Java, or Python. For using any other programming language, please get prior approval from the TA.
 - A ReadMe file with instructions on how to compile/run the code.
 - For all other questions, a PDF/Doc/PS/Image file with the solutions.
- Late Submission Penalty:
 - up to 2 hours late — 10% deduction
 - 2 - 4 hours late — 20% deduction
 - 4 - 12 hours late — 35% deduction
 - 12 - 24 hours late — 50% deduction
 - 24 - 48 hours late — 75% deduction
 - more than 48 hours late — 100% deduction (zero credit)

B. Problems:

1. POS Tagging Errors (10 points)

Find one tagging error in each of the following sentences that are tagged with the Penn Treebank POS tagset in the text book's Figure 5.6 (shown below for your convenience):

1. I/PRP need/VBP a/DT flight/NN from/IN Atlanta/NN
2. Does/VBZ this/DT flight/NN serve/VB dinner/NNS
3. I/PRP have/VB a/DT friend/NN living/VBG in/IN Denver/NNP
4. Can/VBP you/PRP list/VB the/DT nonstop/JJ afternoon/NN flights/NNS

Figure 5.6

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	<i>and, but, or</i>	SYM	symbol	+%, &
CD	cardinal number	<i>one, two, three</i>	TO	“to”	<i>to</i>
DT	determiner	<i>a, the</i>	UH	interjection	<i>ah, oops</i>
EX	existential ‘there’	<i>there</i>	VB	verb, base form	<i>eat</i>
FW	foreign word	<i>mea culpa</i>	VBD	verb, past tense	<i>ate</i>
IN	preposition/sub-conj	<i>of, in, by</i>	VBG	verb, gerund	<i>eating</i>
JJ	adjective	<i>yellow</i>	VBN	verb, past participle	<i>eaten</i>
JJR	adj., comparative	<i>bigger</i>	VBP	verb, non-3sg pres	<i>eat</i>
JJS	adj., superlative	<i>wildest</i>	VBZ	verb, 3sg pres	<i>eats</i>
LS	list item marker	<i>1, 2, One</i>	WDT	wh-determiner	<i>which, that</i>
MD	modal	<i>can, should</i>	WP	wh-pronoun	<i>what, who</i>
NN	noun, sing. or mass	<i>llama</i>	WP\$	possessive wh-	<i>whose</i>
NNS	noun, plural	<i>llamas</i>	WRB	wh-adverb	<i>how, where</i>
NNP	proper noun, singular	<i>IBM</i>	\$	dollar sign	\$
NNPS	proper noun, plural	<i>Carolinas</i>	#	pound sign	#
PDT	predeterminer	<i>all, both</i>	“	left quote	‘ or “
POS	possessive ending	<i>'s</i>	”	right quote	’ or ”
PRP	personal pronoun	<i>I, you, he</i>	(left parenthesis	[, (, {, <
PRP\$	possessive pronoun	<i>your, one's</i>)	right parenthesis],), }, >
RB	adverb	<i>quickly, never</i>	,	comma	,
RBR	adverb, comparative	<i>faster</i>	.	sentence-final punc	. ! ?
RBS	adverb, superlative	<i>fastest</i>	:	mid-sentence punc	: ; ... --
RP	particle	<i>up, off</i>			

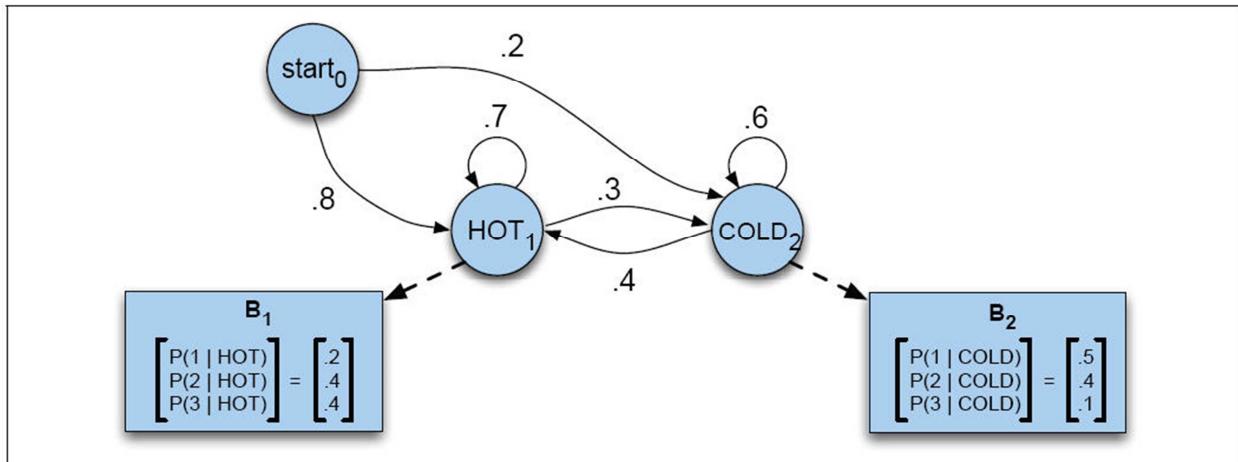
2. Rule Based POS Tagging (30 points)

For this question, you have been given a POS-tagged training file, *HW2_F15_NLP6320_POSTaggedTrainingSet.txt*, that has been tagged with POS tags from the Penn Treebank POS tagset (Figure 5.6). Use this POS tagged file to compute for each word w the tag t that maximizes $P(t|w)$. Retag the training file with POS tags that are most probable for a given word. Compute the error rate by comparing the retagged file against the original tagged file. Now perform error analysis to find the top-5 erroneously tagged words. Write at least five rules to do a better job of tagging these top-5 erroneously tagged words, and show the difference in error rates.

3. HMM Decoding: Viterbi Algorithm (30 points):

Programmatically implement the Viterbi algorithm and run it with the HMM in Fig. 6.3 (shown below for your convenience) to compute the most likely weather sequences for each of the two observation sequences, 331122313 and 331123312.

Figure 6.3



4. Parse Trees (30 points):

Draw tree structures for the following sentences:

1. Does American Airlines have a flight between five a.m. and six a.m.?
2. I would like to fly on American airlines.
3. Please repeat that.
4. Does American 487 have a first-class section?
5. I need to fly between Philadelphia and Atlanta.
6. What is the fare from Atlanta to Denver?

You can use the CFG rules from the Formal Grammars lecture and Chapter 12. More specifically, you can start from the sample grammar rules specified in the text book's Figure 12.10 and Slide 318 (both CFG tables also listed below for your convenience). You can then use the information specified in Section 12.3 ("Some Grammar Rules for English") to create more rules if required for building the parse trees for the example sentences.

Figure 12.10

Grammar	Lexicon
$S \rightarrow NP VP .$	$PRP \rightarrow we he$
$S \rightarrow NP VP$	$DT \rightarrow the that those$
$S \rightarrow "S", NP VP .$	$JJ \rightarrow cold empty full$
$S \rightarrow -NONE-$	$NN \rightarrow sky fire light flight tomorrow$
$NP \rightarrow DT NN$	$NNS \rightarrow assets$
$NP \rightarrow DT NNS$	$CC \rightarrow and$
$NP \rightarrow NN CC NN$	$IN \rightarrow of at until on$
$NP \rightarrow CD RB$	$CD \rightarrow eleven$
$NP \rightarrow DT JJ, JJ NN$	$RB \rightarrow a.m.$
$NP \rightarrow PRP$	$VB \rightarrow arrive have wait$
$NP \rightarrow -NONE-$	$VBD \rightarrow was said$
$VP \rightarrow MD VP$	$VBP \rightarrow have$
$VP \rightarrow VBD ADJP$	$VBN \rightarrow collected$
$VP \rightarrow VBD S$	$MD \rightarrow should would$
$VP \rightarrow VBN PP$	$TO \rightarrow to$
$VP \rightarrow VB S$	
$VP \rightarrow VB SBAR$	
$VP \rightarrow VBP VP$	
$VP \rightarrow VBN PP$	
$VP \rightarrow TO VP$	
$SBAR \rightarrow IN S$	
$ADJP \rightarrow JJ PP$	
$PP \rightarrow IN NP$	

Slide 318

$S \rightarrow NP VP$
$S \rightarrow Aux NP VP$
$S \rightarrow VP$
$NP \rightarrow Pronoun$
$NP \rightarrow Proper-Noun$
$NP \rightarrow Det Nominal$
$Nominal \rightarrow Noun$
$Nominal \rightarrow Nominal Noun$
$Nominal \rightarrow Nominal PP$
$VP \rightarrow Verb$
$VP \rightarrow Verb NP$
$VP \rightarrow Verb NP PP$
$VP \rightarrow Verb PP$
$VP \rightarrow VP PP$
$PP \rightarrow Preposition NP$