NATIONAL UNIVERSITY OF SINGAPORE
SCHOOL OF COMPUTING

CS4248 – Natural Language Processing

Semester 1 AY2017/2018

December 2017 **Time Allowed: 2 Hours**

INSTRUCTIONS TO CANDIDATES

1. This assessment paper contains **EIGHT (8)** questions and comprises **ELEVEN (11)** printed pages, including this page.
2. Answer **ALL** questions within the space in this booklet.
3. This is a **CLOSED** book assessment, but one double-sided A4 sized sheet is allowed for notes.
4. A non-programmable calculator is permitted.
5. Please write your Student Number below:

This portion is for lecturer's use only

Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Total
Max	15	10	20	10	10	10	15	10	100
Marks									

1. (15 marks) Consider the task of word sense disambiguation of the noun *interest*, using a naïve Bayes classifier. Suppose the following 4 training sentences have been collected. Each occurrence of *interest* has been annotated with one of 2 senses s_1 and s_2, as shown below:

s_1: his *interest* is in music
s_1: he has *interest* in mathematics
s_2: the *interest* rate is rising
s_2: the *interest* payment is compounded

Suppose we use surrounding single words (unigrams) as features and generate one training example per sentence. Each training example is a binary feature vector, and each feature has value 0 or 1, where value 0 indicates the absence of the corresponding surrounding word in the training sentence, and value 1 indicates the presence of the word. For example, the feature corresponding to the surrounding word "compounded" has value 1 in the fourth training example, but value 0 in the first three training examples.

Let P_{MLE} denote the maximum likelihood (unsmoothed) probability estimate, P_{Addone} denote the add-one smoothed probability estimate, and P_{WB} denote the Witten-Bell smoothed probability estimate. Fill in the probability estimates indicated in the following table. (These estimates are needed in building a naïve Bayes classifier.)

P_{MLE}(compounded = 0 | s1)	
P_{MLE}(compounded = 1 | s1)	
P_{Addone}(compounded = 0 | s1)	
P_{Addone}(compounded = 1 | s1)	
P_{WB}(compounded = 0 | s1)	
P_{WB}(compounded = 1 | s1)	

Show how you compute the probability estimates and provide the necessary justification in the space below.

(Additional space for answering question 1, if needed)

2. (10 marks) Give a trace of the minimum edit distance algorithm (a dynamic programming algorithm) to compute the minimum cost of transforming the string "sandy" to "wind", by filling out every cell entry in the following table, where each cell entry denotes the minimum cost of transforming the associated substrings. Assume that the cost of inserting a character is 1, the cost of deleting a character is 1, and the cost of substituting a character by a different character is 2. (You do not need to show the optimal path.)

y	5				
d	4				
n	3				
a	2				
s	1				
	0	1	2	3	4
		w	i	n	d

3. (20 marks) Assign one part-of-speech (POS) tag to each word in bold in the following 10 sentences, using the Penn Treebank tagset. Write the POS tag next to each word in bold in the table below.

(a) **To** err is human .
(b) To **err** is human .
(c) **John** moved the chairs to the room .
(d) John **moved** the chairs to the room .
(e) John moved the chairs to **the** room .
(f) John moved the **chairs** to the room .
(g) He is **happier** now .
(h) **There** are apples on the table .
(i) Please pass me the **blue** pen .
(j) He was on leave in November **and** December .

	word	POS tag		word	POS tag
(a)	**To**		(f)	**chairs**	
(b)	**err**		(g)	**happier**	
(c)	**John**		(h)	**There**	
(d)	**moved**		(i)	**blue**	
(e)	**the**		(j)	**and**	

The 45 POS tags in the Penn Treebank tagset are:

CC (Coordin. Conjunction)	PDT (Predeterminer)	VBP (Verb, non-3sg pres)
CD (Cardinal number)	POS (Possessive ending)	VBZ (Verb, 3sg pres)
DT (Determiner)	PRP (Personal pronoun)	WDT (Wh-determiner)
EX (Existential 'there')	PRP$ (Possessive pronoun)	WP (Wh-pronoun)
FW (Foreign word)	RB (Adverb)	WP$ (Possessive wh-)
IN (Preposition/sub-conj)	RBR (Adverb, comparative)	WRB (Wh-adverb)
JJ (Adjective)	RBS (Adverb, superlative)	$ (Dollar sign)
JJR (Adj., comparative)	RP (Particle)	# (Pound sign)
JJS (Adj., superlative)	SYM (Symbol)	" (Left quote)
LS (List item marker)	TO ("to")	" (Right quote)
MD (Modal)	UH (Interjection)	((Left parenthesis)
NN (Noun, sing. or mass)	VB (Verb, base form)	) (Right parenthesis)
NNS (Noun, plural)	VBD (Verb, past tense)	, (Comma)
NNP (Proper noun, singular)	VBG (Verb, gerund)	. (Sentence-final punc)
NNPS (Proper noun, plural)	VBN (Verb, past participle)	: (Mid-sentence punc)

4. Consider the following context-free grammar (CFG):

S → NP VP
NP → PN
VP → V NP
PN → he | him | she | her | we | they | them
V → love | loves | hate | hates

a. (2 marks) Describe the problems with the use of this CFG to model a tiny fragment of English.

b. (8 marks) Give a revised CFG that overcomes the problems.

5. (a) (2 marks) Give one usage of Chomsky normal form in parsing natural language sentences.

5. (b) (8 marks) Convert the following grammar into Chomsky Normal Form (CNF):

$S \rightarrow NP$
$NP \rightarrow N$
$NP \rightarrow NP$ and N
$N \rightarrow$ dog
$N \rightarrow$ (NP)

In this grammar, the set of non-terminal symbols is { *S, NP, N* }, the set of terminal symbols is { and, dog, (,) }

List the productions (grammar rules) of the transformed grammar in CNF in the following box:

Show clearly the steps of your conversion in the space below:

6. (10 marks) A person decides whether to buy a laser printer based on two factors: price and brand. Examples of his past decisions are given below:

Price	Brand	Decision
Expensive	Dell	Yes
Expensive	HP	Yes
Expensive	HP	No
Expensive	HP	No
Medium	Canon	No
Medium	Dell	Yes
Medium	HP	Yes
Medium	HP	Yes
Medium	Samsung	Yes
Cheap	Canon	No
Cheap	Samsung	Yes

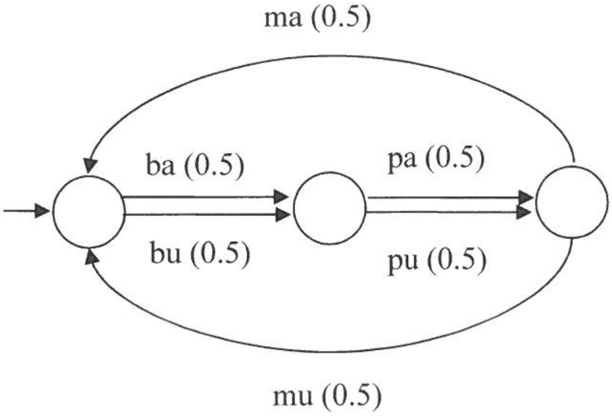
Draw the decision tree that is learned from the above training examples, based on information gain. Show clearly the steps of your calculation.

(Note: $0 \cdot \log_2 0 = 0$, $\log_2 1 = 0$, $\log_2 2 = 1$, $\log_2 3 = 1.585$, $\log_2 4 = 2$, $\log_2 5 = 2.322$,

$\log_2 6 = 2.585$, $\log_2 7 = 2.807$, $\log_2 8 = 3$, $\log_2 9 = 3.170$)

(Additional space for answering question 6)

7. (15 marks) Consider a formal language defined as follows:



That is, the first word is either ba or bu, the second word is either pa or pu, the third word is either ma or mu, the fourth word is either ba or bu, the fifth word is either pa or pu, etc. The transition probability for each word is enclosed in brackets, and the vocabulary of this language is { ba, bu, pa, pu, ma, mu }

Let X be a random variable ranging over all finite sequences of words of length n in this language, with true probability distribution given above.

Consider an incorrect model where the transitions for "ba", "pu", and "ma" are assigned probability of $\frac{2}{5}$, and the transitions for "bu", "pa", and "mu" are assigned probability of $\frac{3}{5}$. What is the (per-word) cross entropy of X using this model? Show clearly the steps of your calculations to justify your answers. **Simplify your answers as much as possible.**

(Additional space for answering question 7, if needed)

8. (a) (5 marks) Give the meaning representation of the following sentence in first order logic (FOL):
"Every pilot who loves a cat is smart."
In the meaning representation, use the following predicates:
P(x): x is a pilot
C(x): x is a cat
S(x): x is smart
L(x, y): x loves y

8. (b) (5 marks) Give the advantages of using FOL for representing the meaning of natural language sentences over the alternative of using frame representation. Be as specific as possible in your answer, which should contain no more than 40 words.

END OF PAPER