Natural Language Processing 1 2016

Assignment 4
Due Thursday 15^{th} Dec.
Total points: 24

1. In the lecture on Distributional Semantics we introduced point wise mutual information (PMI)a measure of statistical independence which can tell use whether two words (or more generally, two statistical events) tend to occur together or not. The PMI between two events x and y is defined as 4 points

$$PMI(x, y) = log_2 \frac{P(x, y)}{P(x)P(y)}$$

Consider two example scenarios:

- x is "eat is the first word in a bigram" and y is "pizza is the second word in a bigram".
- x is "happy occurs in a Tweet" and y is "pizza occurs in a Tweet".

(a) For each example above, say what $P(x, y)$ represents.

(b) What do negative, zero, and positive PMI values represent in terms of the statistical independence of x and y? Give some example pairs of words that you would expect to have negative or positive PMI (in either the bigram or Tweet scenario).

2. Suppose we are using a MaxEnt model for disambiguating three senses of the word *plant*, where y represents the latent sense.

y	=	sense
1		Noun: a member of the plant kingdom
2		Verb: to place in the ground
3		Noun: a factory

(a) We saw the equation for $P(y|\vec{x})$ in a MaxEnt model. Write down a simplified expression for the log probability, $logP(y|\vec{x})$. Can you see why MaxEnt models are also called log-linear models? 2 points

(b) Imagine we have already trained the model. The following table lists the features $\vec{x}$ we are using, and their weights $\vec{w}$ from training: 3 points

feat. #	feature	weight
1	`doc_contains('grow')` & y=1	2.0
2	`doc_contains('grow')` & y=2	1.8
3	`doc_contains('grow')` & y=3	0.3
4	`doc_contains('animal')` & y=1	2.0
5	`doc_contains('animal')` & y=2	0.5
6	`doc_contains('animal')` & y=3	-3.0
7	`doc_contains('industry')` & y=1	-0.1
8	`doc_contains('industry')` & y=2	1.1
9	`doc_contains('industry')` & y=3	2.7

where **`doc_contains('grow')`** means the document containing the target instance of *plant* also contains the word *grow*. Now we see a new document that contains the words industry, grow, and plant. Compute $\sum_i w_i f_i(\vec{x}, y)$ and $P(y|\vec{x})$ for each sense y. Which sense is the most probable?

3. Convert the following FOL expressions into natural language sentences. 3 points

   (a) $\forall x.bear(x) \Rightarrow furry(x)$

   (b) $\exists e.help(e) \wedge helper(e, Jan) \wedge helpee(e, Joost)$

   (c) $\exists e.x.eating(e) \wedge pizza(x) \wedge eater(e, Sergii) \wedge eaten(e, x)$

   (d) $\exists e.x.y.eating(e) \wedge pizza(x) \wedge fork(y) \wedge eater(e, Sergii) \wedge eaten(e, x) \wedge instrument(e, y)$

   (e) $\forall x.student(x) \Rightarrow \exists e.lifting(e) \wedge lifter(e, x) \wedge liftee(e, Marie)$

   (f) $\exists e.lifting(e) \wedge \forall x.student(x) \Rightarrow lifter(e, x) \wedge liftee(e, Marie)$

4. Convert the following natural language sentences into FOL expressions. Use reified event semantics. If a sentence is ambiguous, list all possible interpretations and give paraphrases of the different meanings. 3 points

   (a) Juan hates pasta.

   (b) Some student likes every class.

   (c) Marie sees herself.

5. Assume the following grammar with semantic attachments

$S \rightarrow NP\ VP$	$VP.sem(NP.sem)$
$VP \rightarrow V_t\ NP$	$V_t.sem(NP.sem)$
$VP \rightarrow V_i$	$V_i.sem$
$NP \rightarrow N$	$N.sem$
$N \rightarrow Sam$	Sam
$N \rightarrow Whiskers$	Whiskers
$V_t \rightarrow likes$	$\lambda x.\lambda y.\exists e.Liking(e) \wedge Liker(e, y) \wedge Likee(e, x)$
$V_i \rightarrow meows$	$\lambda x.\exists e.Meowing(e) \wedge Meower(e, x)$

   (a) Show how the meaning of the sentence *Whiskers likes Sam* is built up using this grammar. 3 points

   (b) The FOL MR of *"Whiskers meows"* is $\exists e.Meowing(e) \wedge Meower(Whiskers)$. Can you construct the MR of *"A cat meows"* along the same lines. What is the problem encountered? 2 points

(c) **Extra Credit** 4 points

Suggest a solution. Your solution should give the new grammar with semantic attachment rules, and build the MR for *"A cat meows"*

6. Probabilistic graphical models are a framework to express probabilistic distributions over variables we believe exist in our observations. Typically, a graphical model is built on simplifying assumptions that make our model of reality approachable by statistical learning techniques. You have recently seen one such model, a word alignment model, called IBM model 1. IBM model 1 makes 3 major simplifying assumptions : two of them concern the structure of the graphical model, i.e., the dependency between random variables, and one of them concerns the parameterisation of the model, i.e., the choice of parametric family of a distribution. Discuss all assumptions, and in each case, present an example that shows how the assumption oversimplifies reality. 4 points