11-711: Algorithms for NLP
Homework Assignment #3: Context Free Grammars, CYK Parsing and Semirings

Out: October 22, 2015
Due: November 10, 2015

Problem 1 [25 points]

Recall that a semiring is a 5-tuple $(R, \oplus, \otimes, \bar{0}, \bar{1})$, where R is a set, $\oplus$ and $\otimes$ are binary operations $R \times R \to R$, and $\bar{0}, \bar{1} \in R$, satisfying the semiring laws:

$\forall a, b, c \in R$:

$$(a \oplus b) \oplus c = a \oplus (b \oplus c) \tag{1}$$

$$\bar{0} \oplus a = a \oplus \bar{0} = a \tag{2}$$

$$a \oplus b = b \oplus a \tag{3}$$

$$(a \otimes b) \otimes c = a \otimes (b \otimes c) \tag{4}$$

$$\bar{1} \otimes a = a \otimes \bar{1} = a \tag{5}$$

$$a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c) \tag{6}$$

$$(a \oplus b) \otimes c = (a \otimes c) \oplus (b \otimes c) \tag{7}$$

$$\bar{0} \otimes a = a \otimes \bar{0} = \bar{0}. \tag{8}$$

We showed in class that the Forward algorithm can be used to compute aggregate statistics over all paths from a node s to another node t in a directed acyclic graph. By mapping edge weights into various semirings we can compute various statistics without modifying the algorithm. For instance the semiring

$$(\mathbb{R}_{\geq 0} \cup \{\infty\}, \min, +, \infty, 0)$$

computes the length of the shortest path.

For each of the following quantities of interest, define a semiring which can be plugged into the Forward algorithm to calculate it. Be careful that the object you've defined satisfies the semiring laws (though you are not required to prove it).

1. [**12 points**] Whether there exists *any* path from s to t.

    $R = \{\texttt{True}, \texttt{False}\}$. Define $\oplus$, $\otimes$, $\bar{0}$, and $\bar{1}$.

    How do you map weighted edges into your semiring?

    What is returned when there is no path from s to t?

2. [**13 points**] The set of *all* paths from s to t.

$R = 2^{E^*}$, where E is the set of edges in the graph. Define $\oplus$, $\otimes$, $\bar{0}$, and $\bar{1}$.

How do you map weighted edges into your semiring?

What is returned when there is no path from s to t?

Problem 2 [15 points]

Suppose G is a CFG and we are given two words $w, x \in L(G)$ such that $|x| = 2n$, $|w| = 2n$. We are also given a terminal symbol a and we suppose that our grammar has a rule $A \to a$. What is the *exact* derivation length of xaw in G (that is, the number of derivation steps in which xaw is derived) if:

1. [**7 points**] G is in Greibach Normal Form (GNF)

2. [**8 points**] G is in Chomsky Normal Form (CNF)

Note: we suppose that $xaw \in L(G)$
Explain your answers in both cases.

Problem 3 [20 points]

```
(1) S --> NP VP
(2) VP --> V NP
(3) VP --> V NP PP
(4) NP --> NP NP
(5) NP --> N
(6) PP --> P NP
(7) N --> students
(8) N --> language
(9) N --> instructors
(10) V --> study
(11) N --> study
```

1. [**2 points**] Explain why this grammar is ambiguous.

2. [**18 points**] The grammar given above is not in CNF. Modify it into a CNF grammar that accepts the same language. Your modification should be reversible, meaning that a derivation from your new grammar can be deterministically converted into a derivation from the original grammar. Discuss whether your new grammar is still ambiguous or not.

Problem 4 [40 points]

1. **[20 points]** Given the grammar of the Problem 3, construct a PDA $M = (Q, \Sigma, \Gamma, \delta, q, Z, \emptyset)$ that accepts the language recognized by the grammar, L, by empty stack and has a single state q (i.e. $|Q| = 1$).

2. **[20 points]** Demonstrate that your PDA accepts the same language.