COMS E6998-3, Spring 2011, problem set 2

Due date: 5pm, 6th May 2011

Part 1

In this question we will revisit the problem of non-projective dependency parsing. Given a sentence with n words, a dependency parse y has $y(i,j) = 1$ for any $i \in \{1 \ldots n\}, j \in \{1 \ldots n\}, i \neq j$ if there is a dependency with head word i and modifier word j. The score for any dependency parse is

$$f(y) = \sum_{i,j} \theta(i,j) y(i,j)$$

where $\theta(i,j)$ is the score of dependency (i,j). We define $\mathcal{Y}$ to be the set of non-projective dependency parses; the decoding problem is then to find

$$y^* = \arg\max_{y \in \mathcal{Y}} f(y)$$

This problem can be solved efficiently using algorithms that find maximum-weight directed spanning trees (see the lecture on non-projective parsing).

Now say we define a new set,

$$\mathcal{Z} = \{y : y \in \mathcal{Y}, \;\; y(4,5) = y(10,11)\}$$

Thus $\mathcal{Z}$ is a subset of $\mathcal{Y}$, only including those dependency parses where $y(4,5) = y(10,11)$. The new decoding problem will be to find

$$y^* = \arg\max_{y \in \mathcal{Z}} f(y) \tag{1}$$

In this question, we will derive an algorithm for this problem based on *Lagrangian relaxation*.

Question 1 (10 points)

What is the Lagrangian for the problem in Eq. 1? (Hint: there should be a single Lagrange multiplier, for the constraint $y(4,5) = y(10,11)$.)

Question 2 (10 points)

What is the dual function $L(u)$ for this problem? (Here u is the single Lagrange multiplier.) How can $L(u)$ be calculated efficiently?

Question 3 (25 points)

Give a subgradient algorithm (similar to the algorithm we saw for decoding phrase-based translation systems) for the problem.

Part 2

We are now going to derive a Lagrangian relaxation algorithm for decoding of Markov random fields (MRFs). Consider an MRF over n variables $x_1 x_2 \dots x_n$, where each x_i can be either 0 or 1. Thus there are 2^n possible settings for the variables $x_1 \dots x_n$.

The MRF has an underlying graph with a set of edges E: each edge (member of E) is a pair (i, j) where $i \in \{1 \dots n\}$, $j \in \{1 \dots n\}$, and $i \neq j$.

The score for an assignment to the n variables is then

$$f(x_1, x_2, \dots, x_n) = \sum_{(i,j) \in E} \theta_{i,j}(x_i, x_j)$$

where $\theta_{i,j}(x_i, x_j)$ is a function that scores the pair of values x_i, x_j. The decoding problem is to find

$$\arg \max_{x_1 \dots x_n} f(x_1, x_2, \dots, x_n)$$

We derive the Lagrangian relaxation algorithm as follows. Assume that we have two trees T_1 and T_2 such that $T_1 \subset E$, $T_2 \subset E$, and $T_1 \cup T_2 = E$ (a tree is a set of edges that contains no cycles). Also, assume that we can rewrite

$$f(x_1, x_2, \dots, x_n) = \sum_{(i,j) \in T_1} g_{i,j}(x_i, x_j) + \sum_{(i,j) \in T_2} h_{i,j}(x_i, x_j)$$

where g and h are new functions (it will always be possible to do this, for any definition of the $\theta_{i,j}$ functions).

We can then rewrite the decoding problem as

$$\arg \max_{x_1 \dots x_n, y_1 \dots y_n} \sum_{(i,j) \in T_1} g_{i,j}(x_i, x_j) + \sum_{(i,j) \in T_2} h_{i,j}(y_i, y_j)$$

subject to the constraints

$$x_i = y_i \quad \text{for all } i = 1 \dots n$$

Question 1 (10 points)

What is the Lagrangian for the decoding problem? (Hint: there should be one Lagrange multiplier $u(i)$ for each constraint $x_i = y_i$.)

Question 2 (10 points)

What is the dual function $L(u)$ for this problem? (Here u is the vector of Lagrange multipliers.) How can $L(u)$ be calculated efficiently?

Question 3 (25 points)

Give a subgradient algorithm for the problem.