Reformulation of StructSVM dual in terms of marginal variables over parts  

Simeng Sun

1 Notation

- Let $\mathcal{X}$ be instance space, $y(x_i)$ be the label space of instance $x_i$. The space of all possible labels is thus $\mathcal{Y} = \bigcup_{x_i \in \mathcal{X}} y(x_i)$
- Define loss function $\ell : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$, $\ell(y, y) = 0, \forall y \in \mathcal{Y}$
- Define joint instance label mapping: $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^d$, where $d$ is subject to specific problems to be solved.
- Given training data set $S = ((x^i, y^i))_{i=1}^m$, where each $y^i = (y_1^i, \ldots, y_n^i)$ we can write down the primal problem

2 Primal and Dual problem

The objective function of structural SVM can be written as

$$
\min_w \frac{1}{m} \sum_{i=1}^m \xi_i + \frac{\lambda}{2} \|w\|^2
$$

s.t. $\xi_i \geq \ell(y_i^*, y) - w^T \phi(x_i^*, y) + w^T \phi(x_i, y), \forall i, \forall y \in \mathcal{Y}(x_i)$

The dual problem of structured SVM can then be written as

$$
\max_\alpha \frac{1}{m} \sum_{i=1}^m \sum_{y \in \mathcal{Y}(x_i)} \alpha_{iy} \ell(y^i, y) - \frac{1}{2\lambda m^2} \sum_{i=1}^m \sum_{y \in \mathcal{Y}(x_i)} \sum_{y' \in \mathcal{Y}(x_i)} \alpha_{iy} \alpha_{iy'} \phi_{iy}^T \phi_{iy'}
$$

s.t. $\sum_{y \in \mathcal{Y}(x_i)} \alpha_{iy} = 1, \forall i$

$\alpha_{iy} \geq 0, \forall i, \forall y \in \mathcal{Y}(x_i)$

where $\phi_{iy} = \phi(x_i^i, y^i) - \phi(x_i, y)$

The number of constraints in the primal problem and the number of variables in the dual problem are exponential in the number of labels $n$.

3 Dual Problem via 'parts'

The following reformulation generalizes what we discuss in the lecture which instead of looking at two consecutive words, the 2 words can be at any position. The set of parts of a label $y = (y_1, y_2, \ldots, y_n)$ is thus $R(y) = \{(t, k) : y_t = k\} \cup \{(p, q, k, l) : y_p = k, y_q = l\}$

The loss can thus be decomposed to

$$
\ell(y^i, y) = \sum_{(t, k) \in R(y^i)} l(y^i, (t, k))
$$

$(t, k)$ is any assignment of $y \in \mathcal{Y}(x_i)$ where $t^{th}$ position is $k$. We can do this since the loss count the mismatch of individual word.

The instance label mapping thus becomes $\phi(x^i, (p, q, k, l))$, where $(p, q, k, l)$ is any assignment of $y^i \in \mathcal{Y}(x_i)$ where the $p^{th}$ position is $k$ and the $q^{th}$ position is $l$ for any possible value for $k, l$, for the following $k, l \in [V]$. 
Since \( \sum \alpha_{iy} = 1 \), which can be viewed as a density function over \( y \), the first term of the dual problem can be rewritten to the following form:

\[
\sum_{y \in \mathcal{Y}(x^t)} \alpha_{iy} \ell(y^i, y) = E_{y \sim \alpha_i} \left[ \sum_{(t,k) \in R(y)} \ell(y^i, (t,k)) \right]
\]

Let

\[
\mu_{i,(t,k)} = \sum_{y_{yi=k}} \alpha_{iy}, \forall k, \forall 1 \leq t \leq n^i, \forall i
\]

be the marginal dual variable, which is the probability of \( y \in y(x^t) \) whose \( t^{th} \) position is \( k \). Therefore

\[
E_{y \sim \alpha_i} \left[ \sum_{(t,k) \in R(y)} \ell(y^i, (t,k)) \right] = \sum_{t=1}^{n^i} \sum_{k \in [V]} \mu_{i,(t,k)} \ell(y^i, (t,k))
\]

Similarly for the second term in the dual problem can be written in the form

\[
\sum_{y \in \mathcal{Y}(x^t)} \sum_{y' \in \mathcal{Y}(x^t')} \alpha_{iy} \alpha_{iy'} \phi_{iy} \phi_{iy'} = E_{y \sim \alpha_i} \left[ \sum_{(p,q,k,l) \in R(y)} \phi_{i,y,p,q,k,l}^\top \phi_{j,y',p',q',k',l'} \right]
\]

, where

\[
\phi_{i,y,p,q,k,l} = \phi(x^i) - \phi(x^i, (p,q,k,l))
\]

, where \((p,q,k,l)\) is any assignment of \( y \in \mathcal{Y}(x^t) \) which has \( k \) in the \( p^{th} \) position and \( l \) at \( q^{th} \) position.

Let

\[
\mu_{i,(p,q,k,l)} = \sum_{y_{yi=k} \in l} \alpha_{iy}, \forall k, \forall 1 \leq p \leq n^i, \forall 1 \leq q \leq n^i, \forall i
\]

be marginal dual variable over edge. We need to enforce local consistency between the node and edge marginals so that the dual via ‘parts’ is equivalent to the original dual problem, specifically

\[
\mu_{i,(p,k)} = \sum_{l \in [V]} \mu_{i,(p,q,k,l)}, \forall k, \forall 1 \leq p \leq n^i, \forall 1 \leq q \leq n^i, \forall i
\]

Therefore,

\[
\sum_{y \in \mathcal{Y}(x^t)} \sum_{y' \in \mathcal{Y}(x^t')} \alpha_{iy} \alpha_{iy'} \phi_{iy} \phi_{iy'} = E_{y \sim \alpha_i} \left[ \sum_{(p,q,k,l) \in R(y)} \phi_{i,y,p,q,k,l}^\top \phi_{j,y',p',q',k',l'} \right]
\]

\[
= \sum_{(p,q)} \sum_{(p',q')} \sum_{k \in [V]} \sum_{k' \in [V]} \sum_{l \in [V]} \sum_{l' \in [V]} \mu_{i,(p,q,k,l)} \mu_{j,(p',q',k',l')} \phi_{i,y,p,q,k,l}^\top \phi_{j,y',p',q',k',l'}
\]

where \((p,q)\) indicates every possible pair of positions. Therefore, the original dual problem can be rewritten to the following form in terms of marginal variables \( \mu_{i,(t,k)} \) and \( \mu_{i,(p,q,k,l)} \).

\[
\max_{\mu_{i,(t,k)}, \mu_{i,(p,q,k,l)}} \frac{1}{\lambda m} \sum_{i=1}^{m} \sum_{t=1}^{n^i} \sum_{k \in [V]} \mu_{i,(t,k)} \ell(y^i, (t,k)) - \frac{1}{2\lambda^2 m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{(p,q) \in [V]} \sum_{(p',q') \in [V]} \sum_{k \in [V]} \sum_{k' \in [V]} \sum_{l \in [V]} \sum_{l' \in [V]} \mu_{i,(p,q,k,l)} \mu_{j,(p',q',k',l')} \phi_{i,y,p,q,k,l}^\top \phi_{j,y',p',q',k',l'}
\]

s.t. \( \mu_{i,(p,k)} = \sum_{l \in [V]} \mu_{i,(p,q,k,l)} \), \( \sum_{t=1}^{n^i} \sum_{k \in [V]} \mu_{i,(t,k)} = 1 \)

\[
\mu_{i,(p,q,k,l)} \geq 0 \quad \forall k,l, \forall 1 \leq t \leq n^i, \forall 1 \leq p \leq n^i, \forall 1 \leq q \leq n^i, \forall i
\]

, where

\[
\phi_{i,y,p,q,k,l} = \phi(x^i, y^i) - \phi(x^i, (p,q,k,l))
\]