Supplementary Material for

Training Conditional Random Fields for Maximum Labelwise Accuracy

Samuel S. Gross, Olga Russakovsky, Chuong B. Do, and Serafim Batzoglou

Stanford University, Stanford, CA 94305, USA,
{ssgross,olga,chuongdo,serafim}@cs.stanford.edu

Abstract

In this supplementary material, we derive the recurrences needed for the computation of the $\alpha^\star(i,j)$ and $\beta^\star(i,j)$ matrices in the maximum labelwise accuracy algorithm.

1 Definitions

Recall the definitions of the following matrices from [1]:

$$\alpha(i,j) = \sum_{\mathbf{y}_{1:j}} 1\left\{y_j = i\right\} \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,j}(\mathbf{x},\mathbf{y})\right) \tag{1}$$

$$\beta(i,j) = \sum_{\mathbf{y}_{j:L}} 1\left\{y_j = i\right\} \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{j+1,L}(\mathbf{x},\mathbf{y})\right) \tag{2}$$

$$\alpha^\star(i,j) = \sum_{k=1}^{j} \sum_{\mathbf{y}_{1:j}} 1\left\{y_k = y_k^\star \wedge y_j = i\right\} \cdot Q_k'(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,j}(\mathbf{x},\mathbf{y})\right) \tag{3}$$

$$\beta^\star(i,j) = \sum_{k=j+1}^{L} \sum_{\mathbf{y}_{j:L}} 1\left\{y_k = y_k^\star \wedge y_j = i\right\} \cdot Q_k'(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{j+1,L}(\mathbf{x},\mathbf{y})\right) \tag{4}$$

2 Computing $\alpha^\star(\cdot,\cdot)$

We begin by splitting the inner summation so as to isolate the summations over y_{j-1} and y_j:

$$\alpha^\star(i,j) = \sum_{k=1}^{j} \sum_{\mathbf{y}_{1:j}} 1\left\{y_k = y_k^\star \wedge y_j = i\right\} \cdot Q_k'(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,j}(\mathbf{x},\mathbf{y})\right) \tag{5}$$

$$\begin{aligned} &= \sum_{k=1}^{j} \sum_{\mathbf{y}_{1:j-2}} \sum_{\mathbf{y}_{j-1:j}} 1\left\{y_k = y_k^\star \wedge y_j = i\right\} \cdot Q_k'(\mathbf{w}) \cdot \\ &\qquad \exp\left(\mathbf{w}^T \mathbf{F}_{1,j-1}(\mathbf{x},\mathbf{y})\right) \cdot \exp\left(\mathbf{w}^T \mathbf{f}(y_{j-1}, y_j, \mathbf{x}, j)\right) \end{aligned} \tag{6}$$

$$\begin{aligned} &= \sum_{\mathbf{y}_{j-1:j}} 1\left\{y_j = i\right\} \cdot \exp\left(\mathbf{w}^T \mathbf{f}(y_{j-1}, y_j, \mathbf{x}, j)\right) \cdot \\ &\qquad \sum_{k=1}^{j} \sum_{\mathbf{y}_{1:j-2}} 1\left\{y_k = y_k^\star\right\} \cdot Q_k'(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,j-1}(\mathbf{x},\mathbf{y})\right). \end{aligned} \tag{7}$$

Next, we eliminate the variable y_j by replacing it with i everywhere:

$$
\begin{aligned}
(7) = \sum_{y_{j-1}} \exp\left(\mathbf{w}^T \mathbf{f}(y_{j-1}, i, \mathbf{x}, j)\right) \cdot & \\
\Bigg[1\left\{i = y_j^\star\right\} \cdot Q'_j(\mathbf{w}) \cdot \sum_{\mathbf{y}_{1:j-2}} \exp\left(\mathbf{w}^T \mathbf{F}_{1,j-1}(\mathbf{x}, \mathbf{y})\right) + & \\
\sum_{k=1}^{j-1} \sum_{\mathbf{y}_{1:j-2}} 1\left\{y_k = y_k^\star\right\} \cdot Q'_k(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,j-1}(\mathbf{x}, \mathbf{y})\right) \Bigg]. & \qquad (8)
\end{aligned}
$$

Next, we remove any references to y_{j-1} in the outer summation by replacing it with i'. However, since y_{j-1} is needed in the inner summations, we must now include it there with the constraint $y_{j-1} = i'$:

$$
\begin{aligned}
(8) = \sum_{i'} \exp\left(\mathbf{w}^T \mathbf{f}(i', i, \mathbf{x}, j)\right) \cdot & \\
\Bigg[1\left\{i = y_j^\star\right\} \cdot Q'_j(\mathbf{w}) \cdot \sum_{\mathbf{y}_{1:j-1}} 1\left\{y_{j-1} = i'\right\} \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,j-1}(\mathbf{x}, \mathbf{y})\right) + & \\
\sum_{k=1}^{j-1} \sum_{\mathbf{y}_{1:j-1}} 1\left\{y_k = y_k^\star \wedge y_{j-1} = i'\right\} \cdot Q'_k(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,j-1}(\mathbf{x}, \mathbf{y})\right) \Bigg]. & \qquad (9)
\end{aligned}
$$

Finally, we can substitute the necessary definitions in order to derive a simple recurrence for $\alpha^*(i, j)$:

$$
(9) = \sum_{i'} \exp\left(\mathbf{w}^T \mathbf{f}(i', i, \mathbf{x}, j)\right) \cdot \left[1\left\{i = y_j^\star\right\} \cdot Q'_j(\mathbf{w}) \cdot \alpha(i', j-1) + \alpha^\star(i', j-1)\right]. \qquad (10)
$$

3 Computing $\beta^\star(\cdot, \cdot)$

The derivation for $\beta^\star(i, j)$ is similar; we start by isolating the first two terms of the inner summation:

$$
\begin{aligned}
\beta^\star(i, j) &= \sum_{k=j+1}^{L} \sum_{\mathbf{y}_{j:L}} 1\left\{y_k = y_k^\star \wedge y_j = i\right\} \cdot Q'_k(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{j+1,L}(\mathbf{x}, \mathbf{y})\right) & (11) \\
&= \sum_{k=j+1}^{L} \sum_{\mathbf{y}_{j+2:L}} \sum_{\mathbf{y}_{j:j+1}} 1\left\{y_k = y_k^\star \wedge y_j = i\right\} \cdot Q'_k(\mathbf{w}) \cdot & \\
&\qquad \exp\left(\mathbf{w}^T \mathbf{F}_{j+2,L}(\mathbf{x}, \mathbf{y})\right) \cdot \exp\left(\mathbf{w}^T \mathbf{f}(y_j, y_{j+1}, \mathbf{x}, j+1)\right) & (12) \\
&= \sum_{\mathbf{y}_{j:j+1}} 1\left\{y_j = i\right\} \cdot \exp\left(\mathbf{w}^T \mathbf{f}(y_j, y_{j+1}, \mathbf{x}, j+1)\right) \cdot & \\
&\qquad \sum_{k=j+1}^{L} \sum_{\mathbf{y}_{j+2:L}} 1\left\{y_k = y_k^\star\right\} \cdot Q'_k(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{j+2,L}(\mathbf{x}, \mathbf{y})\right). & (13)
\end{aligned}
$$

Next, we eliminate the y_j variable by replacing it with i:

$$(13) = \sum_{y_{j+1}} \exp\left(\mathbf{w}^T \mathbf{f}(i, y_{j+1}, \mathbf{x}, j+1)\right) \cdot \left[1\left\{y_{j+1} = y^\star_{j+1}\right\} \cdot Q'_{j+1}(\mathbf{w}) \cdot \sum_{\mathbf{y}_{j+2:L}} \exp\left(\mathbf{w}^T \mathbf{F}_{j+2,L}(\mathbf{x}, \mathbf{y})\right) + \sum_{k=j+2}^{L} \sum_{\mathbf{y}_{j+2:L}} 1\left\{y_k = y^\star_k\right\} \cdot Q'_k(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{j+2,L}(\mathbf{x}, \mathbf{y})\right) \right] \tag{14}$$

and replace y_{j+1} with i', changing the scope of the inner summation as before:

$$(14) = \sum_{i'} \exp\left(\mathbf{w}^T \mathbf{f}(i, i', \mathbf{x}, j+1)\right) \cdot \left[1\left\{i' = y^\star_{j+1}\right\} \cdot Q'_{j+1}(\mathbf{w}) \cdot \sum_{\mathbf{y}_{j+1:L}} 1\left\{y_{j+1} = i'\right\} \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{j+2,L}(\mathbf{x}, \mathbf{y})\right) + \sum_{k=j+2}^{L} \sum_{\mathbf{y}_{j+1:L}} 1\left\{y_k = y^\star_k \wedge y_{j+1} = i'\right\} \cdot Q'_k(\mathbf{w}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{j+2,L}(\mathbf{x}, \mathbf{y})\right) \right]. \tag{15}$$

Finally, substituting the necessary definitions, we get

$$(15) = \sum_{i'} \exp\left(\mathbf{w}^T \mathbf{f}(i, i', \mathbf{x}, j+1)\right) \cdot \left[1\left\{i' = y^\star_{j+1}\right\} \cdot Q'_{j+1}(\mathbf{w}) \cdot \beta(i', j+1) + \beta^\star(i', j+1) \right]. \tag{16}$$

4 Computing the difficult term from the gradient

We wish to compute

$$\sum_{k=1}^{L} Q'_k(\mathbf{w}) \cdot \sum_{\mathbf{y}} 1\left\{y_k = y^\star_k\right\} \cdot \mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y}) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y})\right). \tag{17}$$

Using the fact that

$$\mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{L} \sum_{i'} \sum_{i} 1\left\{y_{j-1} = i' \wedge y_j = i\right\} \cdot \mathbf{f}(i', i, \mathbf{x}, j), \tag{18}$$

we obtain

$$(17) = \sum_{k=1}^{L} Q'_k(\mathbf{w}) \cdot \sum_{j=1}^{L} \sum_{i'} \sum_{i} \sum_{\mathbf{y}} 1\left\{y_k = y^\star_k \wedge y_{j-1} = i' \wedge y_j = i\right\} \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y})\right) \tag{19}$$

$$= \sum_{j=1}^{L} \sum_{i'} \sum_{i} \sum_{k=1}^{L} \sum_{\mathbf{y}} Q'_k(\mathbf{w}) \cdot 1\left\{y_k = y^\star_k \wedge y_{j-1} = i' \wedge y_j = i\right\} \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y})\right). \tag{20}$$

The summation over k we can split into three cases: (1) when $k = j$, (2) when $k < j$, and (3) when $k > j$. This gives,

$$
\begin{aligned}
(20) = \sum_{j=1}^{L} \sum_{i'} \sum_{i} \Bigg[& \sum_{\mathbf{y}} Q'_j(\mathbf{w}) \cdot 1\left\{i = y^\star_j \wedge y_{j-1} = i' \wedge y_j = i\right\} \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y})\right) + \\
& \sum_{k=1}^{j-1} \sum_{\mathbf{y}} Q'_k(\mathbf{w}) \cdot 1\left\{y_k = y^\star_k \wedge y_{j-1} = i' \wedge y_j = i\right\} \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y})\right) + \\
& \sum_{k=j+1}^{L} \sum_{\mathbf{y}} Q'_k(\mathbf{w}) \cdot 1\left\{y_k = y^\star_k \wedge y_{j-1} = i' \wedge y_j = i\right\} \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp\left(\mathbf{w}^T \mathbf{F}_{1,L}(\mathbf{x}, \mathbf{y})\right) \Bigg].
\end{aligned}
\tag{21}
$$

Substituting definitions and refactoring gives,

$$
\begin{aligned}
(21) = \sum_{j=1}^{L} \sum_{i'} \sum_{i} \Bigg[& Q'_j(\mathbf{w}) \cdot \alpha(i', j-1) \cdot \beta(i, j) \cdot 1\left\{i = y^*_j\right\} \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp(\mathbf{w}^T \mathbf{f}(i', i, \mathbf{x}, j) + \\
& \alpha^\star(i', j-1) \cdot \beta(i, j) \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp(\mathbf{w}^T \mathbf{f}(i', i, \mathbf{x}, j) + \\
& \alpha(i', j-1) \cdot \beta^\star(i, j) \cdot \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp(\mathbf{w}^T \mathbf{f}(i', i, \mathbf{x}, j) \Bigg]
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
= \sum_{j=1}^{L} \sum_{i'} \sum_{i} & \mathbf{f}(i', i, \mathbf{x}, j) \cdot \exp(\mathbf{w}^T \mathbf{f}(i', i, \mathbf{x}, j) \cdot \\
& \Bigg[Q'_j(\mathbf{w}) \cdot \alpha(i', j-1) \cdot \beta(i, j) \cdot 1\left\{i = y^*_j\right\} + \\
& \alpha^\star(i', j-1) \cdot \beta(i, j) + \alpha(i', j-1) \cdot \beta^\star(i, j) \Bigg].
\end{aligned}
\tag{23}
$$

References

[1] S. S. Gross, O. Russakovsky, C. B. Do, and S. Batzoglou. Training conditional random fields for maximum labelwise accuracy. In *NIPS 19*, 2007.