

Supplemental Material for Entropy and Margin Maximization for Structured Output Learning

Patrick Pletscher, Cheng Soon Ong, and Joachim M. Buhmann

Department of Computer Science, ETH Zürich, Switzerland

A Derivation of the Dual

Here we derive the dual of the regularized risk with the example loss $\ell_\beta(w, x, y)$. The primal Lagrangian (which we have to minimize w.r.t. w) is given by

$$\begin{aligned} \mathcal{L}(w, u) &= \sum_{n=1}^N -\langle w, \phi(x^{(n)}, y^{(n)}) \rangle \\ &\quad + \frac{1}{\beta} \log \sum_y \exp(\beta \langle w, \phi(x^{(n)}, y) \rangle + \beta \Delta(y, y^{(n)})) \\ &\quad + \frac{C}{2} \|w\|_2^2. \end{aligned}$$

We define $z_{n,y} = \langle w, \phi(x^{(n)}, y) \rangle + \Delta(y, y^{(n)})$, we thus get

$$\begin{aligned} \mathcal{L}(w, z, u) &= \sum_{n=1}^N -\langle w, \phi(x^{(n)}, y^{(n)}) \rangle + \frac{1}{\beta} \log \sum_y \exp(\beta z_{n,y}) \\ &\quad + \frac{C}{2} \|w\|_2^2 \\ &\quad - \sum_{n=1}^N \sum_y u_{n,y} [z_{n,y} - \langle w, \phi(x^{(n)}, y) \rangle - \Delta(y, y^{(n)})] \\ &= - \sum_{n=1}^N \sum_y z_{n,y} u_{n,y} + \frac{1}{\beta} \log \sum_y \exp(\beta z_{n,y}) \\ &\quad + \sum_{n=1}^N -\langle w, \phi(x^{(n)}, y^{(n)}) \rangle + \frac{C}{2} \|w\|_2^2 \\ &\quad + \sum_{n=1}^N \sum_y u_{n,y} [\langle w, \phi(x^{(n)}, y) \rangle + \Delta(y, y^{(n)})]. \end{aligned}$$

Taking the derivatives w.r.t. the primal variable w we get

$$\frac{\partial \mathcal{L}(w, z, u)}{\partial w} = \sum_{n=1}^N -\phi(x^{(n)}, y^{(n)}) + Cw + \sum_{n=1}^N \sum_y u_{n,y} \phi(x^{(n)}, y) \stackrel{!}{=} 0,$$

and thus

$$w^* = \frac{1}{C} \sum_{n=1}^N [\phi(x^{(n)}, y^{(n)}) - \sum_y u_{n,y} \phi(x^{(n)}, y)].$$

The dual function is thus

$$\begin{aligned} \inf_{z,w} \mathcal{L}(w, z, u) &= -\frac{1}{2} u^T A u + b^T u \\ &\quad - \sup_z \left[u^T z - \frac{1}{\beta} \sum_{n=1}^N \log \sum_y \exp(\beta z_{n,y}) \right] \\ &= -\frac{1}{2C} u^T A u + b^T u - \frac{1}{\beta} \sum_{n=1}^N \sum_y u_{n,y} \log u_{n,y}. \end{aligned}$$

In addition, we get the constraints $u_{n,y} \geq 0$ and $\sum_{n,y} u_{n,y} = 1$. The last step follows from the minimization w.r.t. z :

$$u_{n,y} = \frac{\exp(\beta z_{n,y})}{\sum_{y'} \exp(\beta z_{y'})}.$$

This is solvable for $u_{n,y} \geq 0$, $\sum_{n,y} u_{n,y} = 1$:

$$z_{n,y}^* = \frac{1}{\beta} \log(u_{n,y}).$$

Putting this back into the primal we get

$$\sum_{n=1}^N \frac{1}{\beta} u_{n,y} \log(u_{n,y}) - \frac{1}{\beta} \sum_{n=1}^N \log(\sum_y \exp(\log(u_{n,y}))) = \frac{1}{\beta} \sum_{n=1}^N \sum_y u_{n,y} \log u_{n,y}.$$

The dual problem is thus:

$$\begin{aligned} \min_u \quad & \frac{1}{2C} u^T A u - b^T u + \frac{1}{\beta} \sum_{n=1}^N \sum_y u_{n,y} \log u_{n,y} \\ \text{s.t.} \quad & u_{n,y} \geq 0 \\ & \sum_y u_{n,y} = 1 \quad \forall n \end{aligned}$$

B Inverse Temperature in the CRF objective

Our goal here is to show that the loss

$$\sum_{n=1}^N -\frac{1}{\beta} \log P_\beta(y^{(n)} | x^{(n)}, w) + \frac{C}{2} \|w\|^2,$$

is equivalent to the standard log-loss with regularizer parameter $C' = C/\beta$.

$$\begin{aligned}
& \sum_{n=1}^N -\frac{1}{\beta} \log P_{\beta}(y^{(n)}|x^{(n)}, w) + \frac{C}{2} \|w\|^2 \\
& \Leftrightarrow \sum_{n=1}^N -\log P_{\beta}(y^{(n)}|x^{(n)}, w) + \frac{\beta C}{2} \|w\|^2 \\
& \Leftrightarrow \sum_{n=1}^N -\log P(y^{(n)}|x^{(n)}, w) + \frac{\beta C}{2} \left\| \frac{w}{\beta} \right\|^2 \\
& \Leftrightarrow \sum_{n=1}^N -\log P(y^{(n)}|x^{(n)}, w) + \frac{C}{2\beta} \|w\|^2.
\end{aligned}$$

And thus $C' = C/\beta$.