Proof: Discriminative Rank Pooling for Activity Recognition

Basura Fernando, Peter Anderson, Marcus Hutter, Stephen Gould
The Australian National University
Canberra, Australia
firstname.lastname@anu.edu.au

In this supplementary note we provide the derivation of Equation (12) in the main paper for updating the parameters $W \in \mathbb{R}^{D \times D}$ of our non-linear function $\psi(W\boldsymbol{x}_t)$. Let us remind the reader that we are given a function of the form

$$\boldsymbol{u}^\star(W) = \operatorname*{argmin}_{\boldsymbol{u}} \left\{ \frac{1}{2}\|\boldsymbol{u}\|^2 + \frac{C}{2}\sum_{t=1}^{J}\left[|t - \boldsymbol{u}^T\psi(W\boldsymbol{x}_t)| - \epsilon\right]^2_{\geq 0} \right\} \tag{1}$$

which we need to differentiate (with respect to W). We start with a lemma.

Lemma 1: Let $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ be a conitunous function with first and second derivatives. Let $\boldsymbol{g}(x) = \operatorname{argmin}_{\boldsymbol{y} \in \mathbb{R}^n} f(x, \boldsymbol{y})$. Then

$$\boldsymbol{g}'(x) = -f_{YY}(x, \boldsymbol{g}(x))^{-1} f_{XY}(x, \boldsymbol{g}(x)).$$

where $f_{YY} \doteq \nabla^2_{\boldsymbol{yy}} f(x, \boldsymbol{y}) \in \mathbb{R}^{n \times n}$ and $f_{XY} \doteq \frac{d}{dx}\nabla_{\boldsymbol{y}} f(x, \boldsymbol{y}) \in \mathbb{R}^n$.

Proof. We have:

$$f_Y(x, \boldsymbol{g}(x)) \doteq \nabla_Y f(x, \boldsymbol{y})|_{\boldsymbol{y}=\boldsymbol{g}(x)} = 0 \tag{2}$$

$$\frac{d}{dx} f_Y(x, \boldsymbol{g}(x)) = 0 \tag{3}$$

$$\therefore f_{XY}(x, \boldsymbol{g}(x)) + f_{YY}(x, \boldsymbol{g}(x))\boldsymbol{g}'(x) = 0 \tag{4}$$

$$\frac{d}{dx}\boldsymbol{g}(x) = -f_{YY}(x, \boldsymbol{g}(x))^{-1} f_{XY}(x, \boldsymbol{g}(x)) \tag{5}$$

□

Now let

$$f(W, \boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u}\|^2 + \frac{C}{2}\sum_{t=1}^{J}\left[|t - \boldsymbol{u}^T\psi(W\boldsymbol{x}_t)| - \epsilon\right]^2_{\geq 0} \tag{6}$$

Let $\boldsymbol{v}_t = \psi(W\boldsymbol{x}_t)$. Then

$$\nabla_{\boldsymbol{u}} \frac{1}{2}\left[|t - \boldsymbol{u}^T\boldsymbol{v}_t| - \epsilon\right]^2_{\geq 0} = \begin{cases} (\boldsymbol{u}^T\boldsymbol{v}_t - t + \epsilon)\boldsymbol{v}_t, & \text{if } t - \boldsymbol{u}^T\boldsymbol{v}_t \geq \epsilon \\ (\boldsymbol{u}^T\boldsymbol{v}_t - t - \epsilon)\boldsymbol{v}_t, & \text{if } \boldsymbol{u}^T\boldsymbol{v}_t - t \geq \epsilon \\ \boldsymbol{0}, & \text{otherwise} \end{cases} \tag{7}$$

$$\nabla^2_{\boldsymbol{uu}} \frac{1}{2}\left[|t - \boldsymbol{u}^T\boldsymbol{v}_t| - \epsilon\right]^2_{\geq 0} = \begin{cases} \boldsymbol{v}_t\boldsymbol{v}_t^T, & \text{if } |t - \boldsymbol{u}^T\boldsymbol{v}_t| \geq \epsilon \\ \boldsymbol{0}, & \text{otherwise} \end{cases} \tag{8}$$

$$\frac{\partial}{\partial W_{ij}}\nabla_{\boldsymbol{u}} \frac{1}{2}\left[|t - \boldsymbol{u}^T\boldsymbol{v}_t| - \epsilon\right]^2_{\geq 0} = \begin{cases} (\boldsymbol{u}^T\frac{\partial \boldsymbol{v}_t}{\partial W_{ij}})\boldsymbol{v}_t + (\boldsymbol{u}^T\boldsymbol{v}_t - t + \epsilon)\frac{\partial \boldsymbol{v}_t}{\partial W_{ij}}, & \text{if } t - \boldsymbol{u}^T\boldsymbol{v}_t \geq \epsilon \\ (\boldsymbol{u}^T\frac{\partial \boldsymbol{v}_t}{\partial W_{ij}})\boldsymbol{v}_t + (\boldsymbol{u}^T\boldsymbol{v}_t - t - \epsilon)\frac{\partial \boldsymbol{v}_t}{\partial W_{ij}}, & \text{if } \boldsymbol{u}^T\boldsymbol{v}_t - t \geq \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Now, let

$$e_t = \begin{cases} t - \boldsymbol{v}_t^T \boldsymbol{u}^\star - \epsilon, & \text{if } t - \boldsymbol{u}^T \boldsymbol{v}_t \geq \epsilon \\ t - \boldsymbol{v}_t^T \boldsymbol{u}^\star + \epsilon, & \text{if } \boldsymbol{u}^T \boldsymbol{v}_t - t \geq \epsilon \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Then

$$f_{UU}(W, \boldsymbol{u}^\star) = I + C \sum_{t:e_t \neq 0} \boldsymbol{v}_t \boldsymbol{v}_t^T \tag{11}$$

$$\frac{\partial}{\partial W_{ij}} f_U(W, \boldsymbol{u}^\star) = -C \sum_{t:e_t \neq 0} e_t \frac{\partial \boldsymbol{v}_t}{\partial W_{ij}} - \langle \boldsymbol{u}^\star, \frac{\partial \boldsymbol{v}_t}{\partial W_{ij}} \rangle \boldsymbol{v}_t \tag{12}$$

$$\therefore \frac{\partial \boldsymbol{u}^\star(W)}{\partial W_{ij}} = \left(I + C \sum_{t:e_t \neq 0} \boldsymbol{v}_t \boldsymbol{v}_t^T \right)^{-1} \left(C \sum_{t:e_t \neq 0} e_t \frac{\partial \boldsymbol{v}_t}{\partial W_{ij}} - \langle \boldsymbol{u}^\star, \frac{\partial \boldsymbol{v}_t}{\partial W_{ij}} \rangle \boldsymbol{v}_t \right) \tag{13}$$

Now, $\boldsymbol{v}_t = \psi(W \boldsymbol{x}_t)$ where $\psi(\cdot)$ operates element-wise. Therefore,

$$\left(\frac{\partial \boldsymbol{v}_t}{\partial W_{ij}} \right)_k = \begin{cases} \psi'(W \boldsymbol{x}_t)_{[k]} \boldsymbol{x}_{t[j]}, & \text{if } k = i \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Moreover, let us approximate Equation 11 by a diagonal matrix. Then,

$$\frac{\partial \boldsymbol{u}_k^\star(W)}{\partial W_{ij}} = \frac{1}{1 + C \sum_{t:e_t \neq 0} \psi_k^2(W \boldsymbol{x}_t)} C \sum_{t:e_t \neq 0} \left([\![k = i]\!] e_t - \boldsymbol{u}_i^\star \psi_k(W \boldsymbol{x}_t) \right) \psi_i'(W \boldsymbol{x}_t) \boldsymbol{x}_{t[j]} \tag{15}$$

Remember that the score function for class c is given by $h_c = \boldsymbol{w}_c^T \boldsymbol{u}^\star(W)$. Then,

$$\frac{\partial h_c}{\partial W_{ij}} = \frac{\partial h_c}{\partial \boldsymbol{u}^\star} \cdot \frac{\partial \boldsymbol{u}^\star}{\partial W_{ij}} \tag{16}$$

$$= \boldsymbol{w}_c^T \frac{\partial \boldsymbol{u}^\star(W)}{\partial W_{ij}} \tag{17}$$

Now define the scaled classifier parameters $\widehat{\boldsymbol{w}}_c$ by

$$\widehat{\boldsymbol{w}}_{c[i]} = \frac{\boldsymbol{w}_{c[i]}}{1 + C \sum_{t:e_t \neq 0} \psi_{[i]}^2(W \boldsymbol{x}_t)} \tag{18}$$

$$\frac{\partial h_c}{\partial W_{ij}} = \widehat{\boldsymbol{w}}_c^T \left(C \sum_{t:e_t \neq 0} e_t \frac{\partial \boldsymbol{v}_t}{\partial W_{ij}} - \boldsymbol{u}^T \frac{\partial \boldsymbol{v}_t}{\partial W_{ij}} \boldsymbol{v}_t \right) \tag{19}$$

Let $\mathbf{1}$ be the all-ones vector of size n. Then let $K_t = \frac{\partial \boldsymbol{v}_t}{\partial W}$ where $(K_t)_{[ij]} \doteq \frac{\partial \boldsymbol{v}_{t[i]}}{\partial W_{ij}}$. Let $s_t = \widehat{\boldsymbol{w}}_c^T \boldsymbol{v}_t$ be a scalar. Then

$$\frac{\partial h_c}{\partial W_{ij}} = C \sum_{t:e_t \neq 0} e_t (K_t)_{[ij]} \widehat{\boldsymbol{w}}_{c[i]} - s_t (K_t)_{[ij]} \boldsymbol{u}_{[i]}^\star \tag{20}$$

which gives our result

$$\nabla_W h_c = C \sum_{\substack{t=1 \\ e_t \neq 0}}^{J} e_t K_t \odot (\widehat{\boldsymbol{w}}_c \cdot \mathbf{1}^T) - s_t K_t \odot (\boldsymbol{u}^\star \cdot \mathbf{1}^T) \tag{21}$$

where $\odot$ is the Hadamard product.

References