

## Supplementary Material

This is the supplementary material for paper “Boosting with Spatial Regularization”. In this document we present the proofs of two lemmas used in the paper.

The first lemma is used to prove Theorem 1, the convergence of the algorithm when using step size:

$$\tilde{\varepsilon} = \min \left\{ 3 \frac{(W_+ - W_-)}{W_+ + 1.36W_-}, \frac{W_+ - W_- + \gamma_{k'}}{W_+ + W_- + 2\lambda K_{k'k'}}, 1 \right\}. \quad (1)$$

In proof, we defined  $f_1(\varepsilon) = W_-e^\varepsilon - W_+e^{-\varepsilon}$  and  $g_1(\varepsilon) = W_-(1+\varepsilon) - W_+(1-\varepsilon) = (W_+ + W_-)\varepsilon - (W_+ - W_-)$  and want to prove the following lemma:

**Lemma 1.** *If  $0 < \varepsilon \leq \min\{3 \frac{(W_+ - W_-)}{W_+ + 1.36W_-}, 1\}$ , then  $f_1(\varepsilon) - g_1(\varepsilon) \leq 0$ .*

*Proof.* The inequality that we would like to solve is:

$$f_1 - g_1 = W_-(e^\varepsilon - 1 - \varepsilon) - W_+(e^{-\varepsilon} - 1 + \varepsilon) = W_-h_1 - W_+h_2 \leq 0, \quad (2)$$

where  $h_1 = e^\varepsilon - 1 - \varepsilon$ , and  $h_2 = e^{-\varepsilon} - 1 + \varepsilon$ .

By assumption  $0 < \varepsilon \leq 1$ , we have

$$e^\varepsilon = 1 + \varepsilon + \frac{\varepsilon^2}{2!} + \frac{\varepsilon^3}{3!} + O(\varepsilon^4) \leq 1 + \varepsilon + \frac{\varepsilon^2}{2!} + \frac{\varepsilon^3}{3!} + 0.06\varepsilon^3, \quad (3)$$

$$e^{-\varepsilon} = 1 - \varepsilon + \frac{\varepsilon^2}{2!} - \frac{\varepsilon^3}{3!} + O(\varepsilon^4) \geq 1 - \varepsilon + \frac{\varepsilon^2}{2!} - \frac{\varepsilon^3}{3!}. \quad (4)$$

Using these approximations for  $e^\varepsilon$  and  $e^{-\varepsilon}$  in  $h_1$  and  $h_2$  guarantees the inequality (2) to hold true. It then gives  $\varepsilon \leq 3 \frac{(W_+ - W_-)}{W_+ + 1.36W_-}$ .  $\square$

The second lemma is used to prove Theorem 2, the grouping effect. We defined  $\alpha^*$  to be the minimizer of

$$\mathcal{L}_{reg}^{exp}(\mathcal{X}, \mathcal{Y}, \alpha) = \sum_{i=1}^m \exp(-y_i \sum_{j=1}^p \alpha_j h_j(x_i)) + \lambda \beta^T K \beta \quad (5)$$

$$= \sum_{i=1}^m \exp(-y_i \sum_{j=1}^p \alpha_j h_j(x_i)) + \lambda \alpha^T Q^T K Q \alpha. \quad (6)$$

with:  $\beta^* = Q\alpha^*$ ,  $\gamma^* = -2\lambda K\beta^*$ , and the corresponding training instance weight  $w^*$ . We used  $\mathcal{H}_k$  to denote the subset of base classifiers acting on component  $k$ , i.e.,  $\mathcal{H}_k = \{h_j \in \mathcal{H} : s(j) = k\}$ .

**Lemma 2.** For any  $k$ ,  $1 \leq k \leq n$ , we have

$$-\gamma_k^* \geq \max_{h_j \in \mathcal{H}_k} \sum_{i=1}^m y_i h_j(x_i) w_i^*. \quad (7)$$

Moreover, if  $\beta_k^* > 0$ , then the equality holds.

*Proof.* For any base classifier  $h_j$  that is evaluated on voxel  $k$ , it follows from optimality that the derivative as defined in equation

$$-\frac{\partial}{\partial \alpha_{j'}} \mathcal{L}_{reg}^{exp}(\mathcal{X}, \mathcal{Y}, \boldsymbol{\alpha}) = \sum_{i=1}^m y_i h_{j'}(x_i) w_i + \gamma_{k'} \quad (8)$$

must be less or equal to 0 (otherwise we can increase  $\alpha_j^*$  by a small amount to make the loss function smaller), therefore

$$-\gamma_k^* \geq \sum_{i=1}^m y_i h_j(x_i) w_i^*. \quad (9)$$

Take the maximum of the right side gives the best lower bound on  $-\gamma_k^*$ , which is the inequality in lemma. Moreover, if  $\beta_k^* > 0$ , then  $\alpha_j^* > 0$  for some  $j$ . In this case, the derivative must be exactly equal to 0 (otherwise we can either increase or decrease  $\alpha_j^*$  by a small amount to make the loss function smaller). Therefore, in this case we have equality.  $\square$