

---

# Supplementary Material: One-vs-Each Approximation to Softmax for Scalable Estimation of Probabilities

---

**Michalis K. Titsias**  
 Department of Informatics  
 Athens University of Economics and Business  
 mtitsias@aueb.gr

## 1 Proof of Proposition 3

Here we re-state and prove **Proposition 3**.

**Proposition 3.** Assume that  $K = 2$  and we approximate the probabilities  $p(y = 1)$  and  $p(y = 2)$  from (2) with the corresponding Bouchard's bounds given by  $\frac{e^{f_1 - \alpha}}{(1 + e^{f_1 - \alpha})(1 + e^{f_2 - \alpha})}$  and  $\frac{e^{f_2 - \alpha}}{(1 + e^{f_1 - \alpha})(1 + e^{f_2 - \alpha})}$ . These bounds are used to approximate the maximum likelihood solution for  $(f_1, f_2)$  by maximizing the lower bound

$$\mathcal{F}(f_1, f_2, \alpha) = \log \frac{e^{N_1(f_1 - \alpha) + N_2(f_2 - \alpha)}}{[(1 + e^{f_1 - \alpha})(1 + e^{f_2 - \alpha})]^{N_1 + N_2}}, \quad (1)$$

obtained by replacing  $p(y = 1)$  and  $p(y = 2)$  in the exact log likelihood with Bouchard's bounds. Then, the global maximizer of  $\mathcal{F}(f_1, f_2, \alpha)$  is such that

$$\alpha = \frac{f_1 + f_2}{2}, \quad f_k = 2 \log N_k + c, \quad k = 1, 2. \quad (2)$$

*Proof.* The lower bound is written as

$$N_1(f_1 - \alpha) + N_2(f_2 - \alpha) - (N_1 + N_2) [\log(1 + e^{f_1 - \alpha}) + \log(1 + e^{f_2 - \alpha})].$$

We will first maximize this quantity wrt  $\alpha$ . For that it suffices to minimize the upper bound on the following log-sum-exp function

$$\alpha + \log(1 + e^{f_1 - \alpha}) + \log(1 + e^{f_2 - \alpha}),$$

which is a convex function of  $\alpha$ . By taking the derivative wrt  $\alpha$  and setting to zero we obtain the stationary condition

$$\frac{e^{f_1 - \alpha}}{1 + e^{f_1 - \alpha}} + \frac{e^{f_2 - \alpha}}{1 + e^{f_2 - \alpha}} = 1.$$

Clearly, the value of  $\alpha$  that satisfies the condition is  $\alpha = \frac{f_1 + f_2}{2}$ . Now if we substitute this value back into the initial bound we have

$$N_1 \frac{f_1 - f_2}{2} + N_2 \frac{f_2 - f_1}{2} - (N_1 + N_2) \left[ \log(1 + e^{\frac{f_1 - f_2}{2}}) + \log(1 + e^{\frac{f_2 - f_1}{2}}) \right]$$

which is concave wrt  $f_1$  and  $f_2$ . Then, by taking derivatives wrt  $f_1$  and  $f_2$  we obtain the conditions

$$\frac{N_1 - N_2}{2} = \frac{(N_1 + N_2)}{2} \left[ \frac{e^{\frac{f_1 - f_2}{2}}}{1 + e^{\frac{f_1 - f_2}{2}}} - \frac{e^{\frac{f_2 - f_1}{2}}}{1 + e^{\frac{f_2 - f_1}{2}}} \right]$$

$$\frac{N_2 - N_1}{2} = \frac{(N_1 + N_2)}{2} \left[ \frac{e^{\frac{f_2 - f_1}{2}}}{1 + e^{\frac{f_2 - f_1}{2}}} - \frac{e^{\frac{f_1 - f_2}{2}}}{1 + e^{\frac{f_1 - f_2}{2}}} \right]$$

Now we can observe that these conditions are satisfied by  $f_1 = 2 \log N_1 + c$  and  $f_2 = 2 \log N_2 + c$  which gives the global maximizer since  $\mathcal{F}(f_1, f_2, \alpha)$  is concave.  $\square$