Supplementary material for: Entropic Affinities: Properties and Efficient Numerical Computation

Miguel Á. Carreira-Perpiñán and Max Vladymyrov

May 7, 2013

Abstract

This document provides proofs for results in Vladymyrov and Carreira-Perpiñán (2013) ("the paper").

1 Equations for the first and second derivatives of F

In the root-finding algorithms we optimized over the parameter $\alpha = \log\beta$ to avoid dealing with negative parameters and to make the problem more linear. The function that we optimize and its derivatives are as follows:

$$\begin{aligned} F(\mathbf{x},\alpha,K) &= e^{\alpha} m_1 + \log Z - \log K \\ \frac{\partial F(\mathbf{x},\alpha,K)}{\partial\alpha} &= e^{2\alpha}(m_1^2 - m_2) \\ \frac{\partial^2 F(\mathbf{x},\alpha,K)}{\partial\alpha^2} &= 2e^{2\alpha}(m_1^2 - m_2) + e^{3\alpha}(m_3 - 3m_1m_2 + 2m_1^3). \end{aligned}$$

2 Bounds for β

We want bounds β_L^* and β_U^* on β, i.e., satisfying $H(\beta_L^*) > \log K > H(\beta_U^*)$, that are easy to compute and as tight as possible. Below, we prove the bounds in the paper.

Given a query point i, call the distances to the nearest neighbors $d_1^2 < d_2^2 < \cdots < d_N^2$. Define $\overline{d^2} = \frac{1}{N}\sum_{n=1}^N d_n^2$, $\overline{d^4} = \frac{1}{N}\sum_{n=1}^N d_n^4$, $\Delta_N^2 = d_N^2 - d_1^2$, $\Delta_2^2 = d_2^2 - d_1^2$. (If some of the distances are equal, the results hold taking Δ_2^2 as the first nonzero $d_n^2 - d_1^2$.)

Lemma 2.1. *The individual terms of the entropy (eq. (3) in the paper) can be bounded as follows:*

$$d_1^2 < \sum_{n=1}^N p_n d_n^2 < \overline{d^2} \tag{1}$$

$$N\exp(-\overline{d^2}\beta) < \sum_{n=1}^N \exp(-d_n^2\beta) < N\exp(-d_1^2\beta) \tag{2}$$

$$\log N - \overline{d^2}\beta < \log\Big(\sum_{n=1}^N \exp(-d_n^2\beta)\Big) < \log N - d_1^2\beta \tag{3}$$

$$N\overline{d^2}\exp\big(-\tfrac{\overline{d^4}}{\overline{d^2}}\beta\big) < \textstyle\sum_{n=1}^N d_n^2\exp(-d_n^2\beta) \tag{4}$$

$$\frac{1}{1+(N-1)\exp(-\Delta_2^2\beta)} < p_1(\beta) < \frac{1}{1+(N-1)\exp(-\Delta_N^2\beta)}. \tag{5}$$

Proof. The first inequality in (1) comes trivially from the fact that the distances are sorted. For the second one note that $g(\beta) = \sum_{n=1}^N p_n d_n^2$ is strictly monotonically increasing, since $g'(\beta) = \mathrm{E}_p\left\{d_n^4\right\} - \left(\mathrm{E}_p\left\{d_n^2\right\}\right)^2 = \mathrm{var}_p\left\{d_n^2\right\} > 0$ and $\lim_{\beta\to 0} g(\beta) = \overline{d^2}$. The first inequality in (2) can be obtained using Jensen's inequality with coefficients $\frac{1}{N}$. The second one can be derived easily using the fact they the distances are sorted. Inequality (3) comes from (2) by exponentiation. Inequality (4) is obtained using Jensen's inequality with coefficients $1/\sum_{n=1}^N d_n^2$. Finally, (5) is obtained by dividing by d_1^2 and replacing d_n^2 for $n = 2,\dots,N$ with d_n^2 (where n is the smallest integer such that $d_n^2 \neq d_1$) or d_N^2, respectively. $\square$

Lemma 2.2. *The function* $h(x) = -2(1-x)\log\frac{2(1-x)}{N}$ *is strictly monotonically decreasing for* $x \in [3/4, 1]$

Proof. Consider the function $f(x) = -x\log\frac{x}{N}$. This has a positive derivative $f'(x) = -(1+\log\frac{x}{N})$ in the interval $x \in (0, \frac{N}{e})$. Then, for $N > \frac{e}{2}$ (so for all $N > 1$ if N is integer), $f(x)$ is strictly monotonically increasing in the interval $x \in (0, \frac{1}{2}]$. Thus, $h(x) = f\big(2(1-x)\big)$ is strictly monotonically decreasing in the interval $p_1 \in [3/4, 1]$. □

Theorem 2.3 (Cover and Thomas, 1991, Th. 16.3.2)**.** *Let* P *and* Q *be two probability mass functions such that* $\|P-Q\|_1 \leq \frac{1}{2}$*. Then*

$$|H(P) - H(Q)| \leq -\left\|P-Q\right\|_1 \log\frac{\|P-Q\|_1}{N}.$$

Now we have everything ready to prove Theorem 1.1 from the paper.

Theorem 2.4. *The lower and the upper bounds for* β *can be found using the formulae:*

$$\beta_L = \max\left(\frac{N}{N-1}\frac{\log\frac{N}{K}}{\Delta_N^2}, \sqrt{\frac{\log\frac{N}{K}}{d_N^4 - d_1^4}}\right) \tag{6}$$

$$\beta_U = \frac{1}{\Delta_2^2}\log\Big(\frac{p_1}{1-p_1}(N-1)\Big) \tag{7}$$

where p_1 *is the only solution in the interval* $[3/4, 1]$ *of the equation:*

$$2(1-p_1)\log\frac{N}{2(1-p_1)} = \log\big(\min(\sqrt{2N}, K)\big). \tag{8}$$

Proof. Let us prove the lower bound first. Using (3) we can bound the second term in the entropy with

$$H(\beta) > \beta\sum_{n=1}^{N} p_n d_n^2 + \log N - \beta\overline{d^2}.$$

Now applying inequalities (1) and (4) separately on the first term above and using the fact that $\exp(x) \geq 1+x$ we get the following two bounds:

$$H(\beta) > \log N - \beta(\overline{d^2} - d_1^2) = \log N - \beta\left(\frac{1}{N}\sum_{n=2}^{N} d_n^2 + (1-\frac{1}{N})d_1^2\right) > \log N - \beta(1-\frac{1}{N})\Delta_N^2$$

$$\begin{aligned} H(\beta) &> \beta\frac{N\overline{d^2}\exp(-\beta\frac{\overline{d^4}}{\overline{d^2}})}{N\exp(-\beta d_1)} + \log N - \beta\overline{d^2} = \log N - \beta\overline{d^2}\big(1-\exp(-\beta(\tfrac{\overline{d^4}}{\overline{d^2}} - d_1^2))\big) \\ &\geq \log N - \beta^2(\overline{d^4} - d_1^2\overline{d^2}) \geq \log N - \beta^2(d_N^4 - d_1^4). \end{aligned}$$

Equating the right part of both inequalities to $\log K$ and solving for β we get $\beta_{L1} = \frac{N}{N-1}\frac{\log N - \log K}{\Delta_N^2}$ and $\beta_{L2} = \sqrt{\log\frac{N}{K}/(d_N^4 - d_1^4)}$. The maximum of those values will give the desired lower bound.

Now the upper bound. Using theorem 2.3 and choosing $Q_n = \delta_{n1},\ n = 1, \dots, N$ we have $\left\|P-Q\right\|_1 = |p_1 - 1| + \sum_{n=2}^{N}|p_n| = 2(1-p_1)$. We then have that if $p_1(\beta) \geq \frac{3}{4}$ then $H(\beta) \leq h(p_1(\beta))$, where $h(p_1) = -2(1-p_1)\log\frac{2(1-p_1)}{N}$. Thus, if we can show that $h(p_1) \leq \log K$ for all $p_1 \in [\frac{3}{4}, 1]$, then $H(\beta) \leq \log K$ and β will be the required β_U.

First of all, using lemma 2.2 we can turn the condition $p_1 \in [\frac{3}{4}, 1]$ into $h(p_1) \in (0, \log\sqrt{2N}]$. Thus, to satisfy both conditions above, we need to find β such that $h(p_1(\beta)) \leq \log(\min(\sqrt{2N}, K))$. We can solve for p_1 numerically the equation $h(p_1) = \min(\log\sqrt{2N}, \log K)$ e.g. using Newton's method or a fixed-point iteration (the problem is well defined and has unique solution, from lemma 2.2). However, solving $p_1(\beta)$ for β is still costly, so we can use inequality (5) to bound p_1 with a function $\Pi(\beta) = 1/(1+(N-1)\exp(-\Delta_2^2\beta))$. Finally, we have $H(\beta) \leq h(p_1(\beta)) < h(\Pi(\beta)) = \log(\min(\sqrt{2N}, \log K))$. Solving the last equation for β gives the bound sought. □

References

T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications. John Wiley & Sons, New York, London, Sydney, 1991.

M. Vladymyrov and M. Á. Carreira-Perpiñán. Entropic affinities: Properties and efficient numerical computation. In *Proc. of the 30th Int. Conf. Machine Learning (ICML 2013)*, Atlanta, GA, June 16–21 2013.