Correction to Recurrence Proof for Herding

Yutian Chen & Max Welling
Department of Computer Science
University of California, Irvine
yutianc@ics.uci.edu

Abstract

In [2] and [1] a proof was provided for the statement that the attractor set for the weights of herding would lie within a ball of finite radius. A small bug in that proof was pointed out by Olivier Delalleau. This note is a correction to that proof.

1 Old Proof of "Recurrence"

Below we repeat the proposition:

Proposition 2: $\exists$ radius $\mathcal{R}$ such that an idealized herding update performed outside this radius, will always decrease the norm $||\mathbf{w}||_2$.

and its corollary:

Corollary: $\exists$ radius $\mathcal{R}'$ such that a herding algorithm initialized inside a ball with radius $\mathcal{R}'$ will never generate weights $\mathbf{w}$ with norm $||\mathbf{w}||_2 > \mathcal{R}'$.

We first recall the following lemma:

Lemma 1: If $|g_\alpha(s_\alpha)| < \infty, \ \forall \mathbf{s}, \alpha$, then $\exists\, \mathcal{B}$ such that $||\nabla \ell_0||_2 < \mathcal{B}$.

In the following section, we describe the corrected proof of proposition 2.

2 Correction to Proof

We will use the following three facts:

1. $\sum_\alpha w_\alpha \nabla_{w_\alpha} \ell_0 = \ell_0 < 0$ outside the origin,
2. $\mathcal{B}$ is constant,
3. $\ell_0(\beta \mathbf{w}) = \beta \ell_0(\mathbf{w})$

These properties are only sufficient to prove that in any direction of $\mathbf{w}$, represented by $\mathbf{u}$, $\|\mathbf{u}\| = 1$, there is a radius $\mathcal{R}(\mathbf{u})$ beyond which the norm of $\mathbf{w}$ will decrease when $\mathbf{w}$ is in that direction, (i.e. in math $\delta\|w\|_2 < 0, \forall \mathbf{w} = r\mathbf{u}, r > \mathcal{R}$). But we can't claim that there exists a constant $\mathcal{R}$ irrelevant to $\mathbf{u}$ beyond which the norm of any $\mathbf{w}$ will decrease. The corrected proof is as follows:

Write the herding update as $w'_\alpha = w_\alpha + \nabla_{w_\alpha} \ell_0$. Take the inner product with w'_α leading to, $\|\mathbf{w}'\|_2^2 = \|\mathbf{w}\|_2^2 + 2\sum_\alpha w_\alpha \nabla_{w_\alpha} \ell_0 + \|\nabla_{w_\alpha} \ell_0\|_2^2$, which leads to $\delta\|\mathbf{w}\|_2^2 < 2\ell_0 + \mathcal{B}^2$.

Denote the unit hypersphere as $U = \{\mathbf{w} | \|\mathbf{w}\|_2^2 = 1\}$. Since ℓ_0 is continuous on U, and U is a bounded closed set, ℓ_0 can achieve its supremum on U, that is, we can find a maximum point $\mathbf{w}^*$ on U where $\ell_0(\mathbf{w}^*) \geq \ell_0(\mathbf{w}), \forall \mathbf{w} \in U$.

Now combining this with fact 1, the maximum of ℓ_0 on U is negative. And taking into account fact 2, there is some radius $\mathcal{R}$ for which $\mathcal{R}\ell_0(\mathbf{w}^*) < -\mathcal{B}^2/2$. Together with the scaling property of ℓ_0

from fact 3, we can now prove that any ℓ_0 with a norm larger than $\mathcal{R}$ is smaller than $-\mathcal{B}^2/2$:

$$\ell_0(\mathbf{w}) = \|\mathbf{w}\|_2 \ell_0(\mathbf{w}/\|\mathbf{w}\|_2) \leq \mathcal{R}\ell_0(\mathbf{w}^*) < -\mathcal{B}^2/2, \forall \|\mathbf{w}\| > \mathcal{R} \tag{1}$$
$$\tag{2}$$

Thus, the norm of $\mathbf{w}$ will decrease when it's beyond the bound of $\mathcal{R}$ because $\delta\|\mathbf{w}\|_2^2 < 2\ell_0(\mathbf{w}) + \mathcal{B}^2 < 0, \forall \|\mathbf{w}\|_2 > \mathcal{R}$.

References

[1] M. Welling. Herding dynamic weights for partially observed random field models. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence*, Montreal, Quebec, CAN, 2009.

[2] M. Welling. Herding dynamical weights to learn. In *Proceedings of the 21st International Conference on Machine Learning*, Montreal, Quebec, CAN, 2009.