Bio-inspired robust machine learning

[1,2,3] Ilona Kulikovskikh, [2] Tomislav Šmuc

[1] Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
Centre of Research Excellence for Data Science and Advanced Cooperative Systems
[2] Division of Electronics, Ruđer Bošković Institute, Zagreb, Croatia
[3] Samara National Research University, Samara, Russia

INTRODUCTION

Recent research in machine learning pointed to the core problem of state-of-the-art models which impedes their widespread adoption in different domains. The models' inability to differentiate between noise and subtle, yet significant variation in data leads to their vulnerability to adversarial perturbations that cause wrong predictions with high confidence.

The study is aimed at identifying **whether the algorithms inspired by biological evolution may achieve better results in cases where brittle robustness properties are highly sensitive to the slight noise.** To answer this question, we introduce a new learning procedure inspired by the stability and adaptability of biological systems to unknown and changing environments. The new optimization technique involves an open-ended adaptation process with regard to two hyperparameters inherited from the generalized Verhulst population growth equation. The hyperparameters increase robustness to adversarial noise by penalizing the degree to which hardly visible changes in gradients impact prediction.

The results of computational experiments showed consistency and confirm the viability of bio-inspired machine learning models. We concluded that they may contribute to better robustness to adversarial noise by penalizing the degree to which subtle changes in gradients impact prediction.

PROBLEM STATEMENT

We consider a dataset $\{\mathrm{x}_i, y_i\}_{i=1}^m$ with $\mathrm{x}_i \in \mathbb{R}^n$, $y_i \in \{0,1\}$ and minimize an empirical loss function:

$$\mathscr{L}(\boldsymbol{\theta}) = \sum_{i=1}^{m} \ell(\boldsymbol{\theta}^T \mathrm{x}_i),$$

with a weight vector $\boldsymbol{\theta} \in \mathbb{R}^n$. We are interested in linearly separable problems with a smooth monotone strictly decreasing and non-negative loss function.

The solution to the problem $\min_{\boldsymbol{\theta}\in\mathbb{R}^n} \mathscr{L}(\boldsymbol{\theta})$ can be found using i^{th} iteration of gradient descent updates with a learning rate η:

$$\boldsymbol{\theta}_{l+1} = \boldsymbol{\theta}_l - \eta \nabla \mathscr{L}(\boldsymbol{\theta}_l) = \boldsymbol{\theta}_l - \eta \sum_{i=1}^{m} \ell'(\boldsymbol{\theta}_l^{\mathrm{T}} \mathrm{x}_i)\mathrm{x}_i.$$

It is assumed that $\forall i \in \{1, \dots, m\}$: $y_i = 1, \|\mathrm{x}_i\| < 1$.

BIO-INSPIRED LOSS FUNCTION

We introduce the bio-inspired logistic loss function $l_r(t; a, b)$ with regard to the generalized Verhulst growth model and its solution:

$$\ell_r'(t; a, b) = r(\ell_r(t; a, b) - a)\left(1 - \frac{\ell_r(t; a, b) - a}{b - a}\right)$$

$$\forall t \in \mathbb{R} : \ell_r'(t; a, b) < 0, \lim_{t\to\infty} \ell_r'(t; a, b) = \lim_{t\to-\infty} \ell_r'(t; a, b) = 0$$

$$\ell_r(t; a, b) = a + \frac{b - a}{\left(1 - \left(1 - \left(\frac{b-a}{P_0 + a}\right)\right)\exp(-rt)\right)},$$

so that $\forall t \in \mathbb{R} : \ell_r(t; a, b) > 0, \lim_{t\to\infty} \ell_r(t; a, b) = b - a, \lim_{t\to-\infty} \ell_r(t; a, b) = a$, $P_0 + a = \frac{b-a}{2}$

where the upper asymptote *b* is the population carrying capacity, the lower asymptote *a* indicates critical population thresholds $0 \le a < b \le 1$ below which a population crashes to extinction.

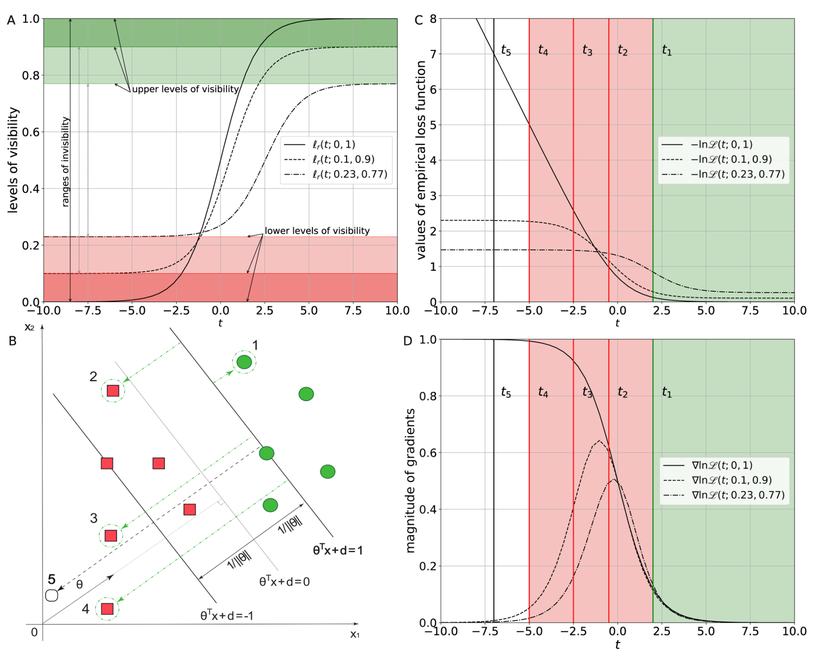

Fig 1. The influence of a difference in approaching the lower and the upper asymptote on the magnitude of gradients.
(A) The definitions of levels of visibility and ranges of invisibility based on the generalized logistic loss function ℓr(t; a, b) subject to the different pairs (a, b) = {(0, 1), (0.1, 0.9), (0.23, 0.77)} and r = 1. (B) A decision boundary with reliable (1), noisy (2,3,4), and adversarial (5) instances. (C) The empirical generalized logistic loss function −ln $\mathscr{L}$(t; a, b) with the levels of visibility for reliable t1, noisy t2, t3, t4, and adversarial t5 instances subject to the different pairs (a, b) = {(0, 1), (0.1, 0.9), (0.23, 0.77)} and r = 1. (D) The magnitude of the gradients of the empirical generalized logistic loss function ∇ln $\mathscr{L}$(t; a, b) subject to the different pairs (a, b) = {(0, 1), (0.1, 0.9), (0.23, 0.77)} and r = 1.

BIO-INSPIRED GRADIENT DESCENT

The empirical generalized logistic loss function and its gradient can be represented as follows:

$$\ln \mathscr{L}(\boldsymbol{\theta}; a, b) = -\sum_{i=1}^{m} y_i \ln \ell_r(\boldsymbol{\theta}^T \mathrm{x}_i; a, b) + (1 - y_i) \ln(1 - \ell_r(\boldsymbol{\theta}^T \mathrm{x}_i; a, b)),$$

$$\nabla \ln \mathscr{L}(\boldsymbol{\theta}; a, b) = -\sum_{i=1}^{m} (y_i - \ell_r(\boldsymbol{\theta}^{\mathrm{T}} \mathrm{x}_i; a, b)) \frac{\ell_r'(\boldsymbol{\theta}^{\mathrm{T}} \mathrm{x}_i; a, b)}{(1 - \ell_r(\boldsymbol{\theta}^{\mathrm{T}} \mathrm{x}_i; a, b))\ell_r(\boldsymbol{\theta}^{\mathrm{T}} \mathrm{x}_i; a, b)} \mathrm{x}_i, \forall y_i \in \{0, 1\}.$$

Taking into account our assumption that $y_i = 1$, the bio-inspired gradient descent can be given as:

$$\boldsymbol{\theta}_{l+1} = \boldsymbol{\theta}_l - \eta \sum_{i=1}^{m} \frac{\ell_r'(\boldsymbol{\theta}_l^{\mathrm{T}} \mathrm{x}_i; a, b)}{\ell_r(\boldsymbol{\theta}_l^{\mathrm{T}} \mathrm{x}_i; a, b)} \mathrm{x}_i = \boldsymbol{\theta}_l - \eta_r \sum_{i=1}^{m} \frac{\ell'(\boldsymbol{\theta}_l^{\mathrm{T}} \mathrm{x}_i)}{\ell_r(\boldsymbol{\theta}_l^{\mathrm{T}} \mathrm{x}_i; a, b)} \mathrm{x}_i.$$

BioGD represents an implementation of bio-inspired gradient descent (see Algorithm).
The algorithm optimizes *a*, *b* and *r* with a grid search to provide more reliable results by the exclusion of unnecessary variance in the estimates of hyperparameters.

Algorithm Bio-inspired gradient descent

procedure BioGD(x, y, η, n)
  Initialize $\boldsymbol{\theta}_0$;
  Initialize $a \in [a_{\min}, a_{\max}]$, $b \in [b_{\min}, b_{\max}]$, $r \in [r_{\min}, r_{\max}]$;
  Initialize a grid of n points in the space $a \times b \times r$;
  Split (x, y) into train $(\mathrm{x}, y)_T$ and cross-validation $(\mathrm{x}, y)_{CV}$ subsets;
  $l \leftarrow 0$;
  repeat
    $\boldsymbol{\theta}_{l+1} \leftarrow \boldsymbol{\theta}_l - (b - a)r\eta\nabla_{(\boldsymbol{\theta})}\mathscr{L}(\boldsymbol{\theta}_l, (\mathrm{x}, y)_T)$;
    $l \leftarrow l + 1$;
  until converge
  $(a, b, r) \leftarrow GridSearch(\mathscr{L}(\boldsymbol{\theta}_{l+1}, (\mathrm{x}, y)_{CV}), a, b, r)$;
  return $\boldsymbol{\theta}_{l+1}, (a, b, r)$

COMPUTATIONAL EXPERIMENTS

Synthetic datasets

We created an n-dimensional dataset so that each dimension's mean μ_j is sampled from a Gaussian distribution $N(0,1)$ while each dimension's standard deviation σ_j is generated according to a distribution $N(1,1)$. The feature space of n-dimensional instances is sampled from the distribution $N(\mu_j, \sigma_j)$. The instances in the dataset were labelled by a randomly chosen hyperplane, so that the generated dataset is linearly separable. For the originally generated dataset, the portion of mislabelled examples is *p* = 0.0. We constructed the noisy versions of this dataset by flipping the labels for the randomly selected proportion of data instances *p* = {0.05, 0.1, 0.15, 0.2, 0.25, 0.3}.

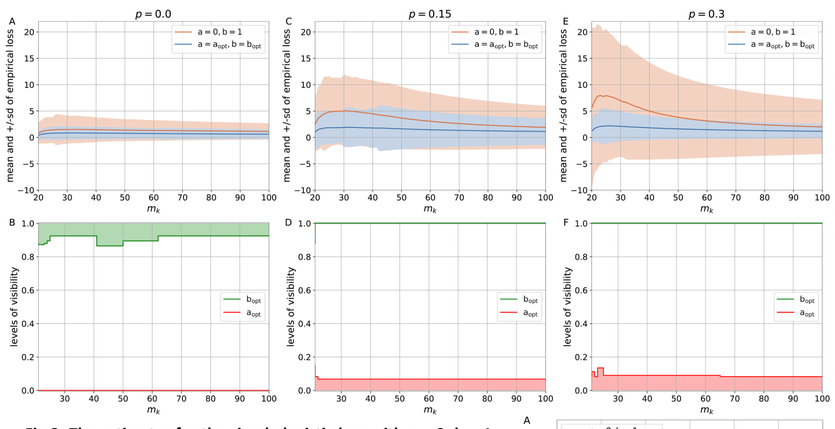

Fig 2. The estimates for the simple logistic loss with a = 0, b = 1 and the generalized logistic loss with the optimal levels of visibility a_{opt}, b_{opt} subject to $m_k \in$ [20, 100] and *p* = {0.0, 0.15, 0.3}.
(A) The mean and +/-sd of empirical loss subject to *p* = 0.0. (B) The levels of visibility a_{opt}, b_{opt} subject to *p* = 0.0. (C) The mean and +/-sd of empirical loss subject to *p* = 0.15. (D) The levels of visibility a_{opt}, b_{opt} subject to *p* = 0.15. (E) The mean and +/-sd of empirical loss subject to *p* = 0.3. (F) The levels of visibility a_{opt}, b_{opt} subject to *p* = 0.3.

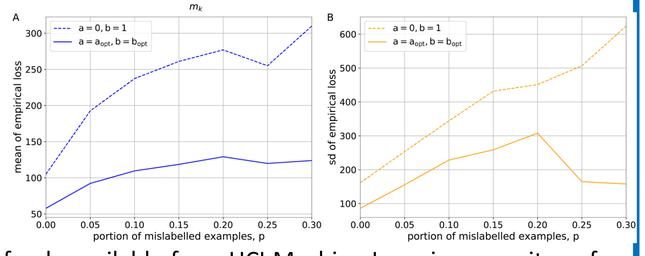

Fig 3. The estimates for the simple logistic loss with a = 0, b = 1 and the generalized logistic loss with the optimal levels of visibility a_{opt}, b_{opt} subject to the proportions of noisy labels *p* = {0.0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3}.
(A) The mean of empirical loss summed over the interval m_k = [20, 100]. (B) The sd of empirical loss summed over the interval m_k = [20, 100].

Experimental datasets

We chose 11 experimental datasets, which are freely available from UCI Machine Learning repository, for validating BioGD. Incrementally varying a number of instances $m_k \in$ [20, 100] randomly taken from the original dataset with number of instances *m*, we explored the adaptivity of the hyperparameters a_{opt} and b_{opt} to the newly updated data.

Fig 4. The estimates for the simple loss and the bio-inspired loss with the optimal levels of visibility a_{opt}, b_{opt} subject to $m_k \in$ [20, 100].
(A) The mean and +/-sd of empirical loss for the *vertebral* dataset. (B) The levels of visibility a_{opt}, for the *vertebral* dataset. (C) The mean and +/-sd of empirical loss for the *liver* dataset. (D) The levels of visibility a_{opt}, b_{opt} for the *liver* dataset.

Fig 5. The differences in mean and sd of empirical loss between the simple loss and the bio-inspired loss for all the datasets summed over $m_k \in$ [20, 100].

Real-world application

The neural network was trained on the freely available MNIST handwritten digits dataset to solve a 10-class classification problem. We tested the model on a subset with adversarial perturbations generated with FGSM.

Fig 7. The probabilities of predicting correct labels in the presence of adversarial perturbations for the bio-inspired sigmoid activation function and the simple sigmoid activation function.
(A) The hyperparameters are equal to a_{opt} = 0.017, b_{opt} = 0.93, r_{opt} = 2.8.
(B) The hyperparameters are equal to $a_{threshold}$ = 0.13, b_{opt} = 0.93, r_{opt} = 2.8.

ACKNOWLEDGMENT

This research was funded by the European Union through the European Regional Development Fund, under the grant KK.01.1.1.01.0009 (DATACROSS), the Ministry of Science and Higher Education of Russian Federation (Russian Federation President grant No. MK-6218.2018.9), and the Russian Foundation for Basic Research (grant No. 18-37-00219).

Europska unija
Zajedno do fondova EU

Ministry of Science and Education

EUROPSKI STRUKTURNI I INVESTICIJSKI FONDOVI

Operativni program KONKURENTNOST I KOHEZIJA

This project was supported by European Union's European Regional Development Fund

Ilona Kulikovskikh
FER, University of Zagreb,
Unska 3, Zagreb, Croatia
ilona@irb.hr