IST 597: Foundations of Deep Learning **(Due: 03/02/22)**

Assignment #00100

Instructor: **Dr. Lee Giles** *TA:* **Ankur Mali** *Name: , PSUID:*

Created by Ankur Mali

Course Policy: Carefully read all the instructions, before you start working on the assignment

- Please typeset your submissions in any LaTeX or word template, give maximum explanation for each sub-problems. Please include your name and PSUID with submission along with date and time on the first page.
- Assignments are due before class at 02:29:59 pm {Please check the due date on the **Official course** webportal} Portal.
- Please avoid single line answers, submissions without any explanations would receive **0** points.
- Late assignments will suffer 50 percent points reductions per day, formula is $x/2$, where x=number of days and counter will start at 02:30:00 pm.
- All source materials must be cited. The University Academic Code of Conduct will be strictly enforced.
- We will be creating Canvas submission page for this. Submit all files on Canvas.
- All queries related to Assignment should have a subject line IST597:Assignment00100 Queries

Problem 1. Build your own custom optimizer. (7 points)

This assignment will build a custom stochastic algorithm to update your model weights. You will modify the starter code provided for assignment one and build on top of it. In other words, you will replace the Keras optimizer with a custom build optimizer (algorithm 1). You will compare the custom optimizer with Keras inbuild optimizers (SGD, RMSProp, and Adam) and show performance across ten trials. Report your findings and comment on speed, stability, and robustness. Note:- Based on assignment 1, select the best model (with and without regularization) for each dataset. You should have total 12 models for both datasets.

Algorithm: Stochastic optimization. Here g_t^2 indicates the element-wise square $g_t \odot g_t$ and g_t^3 indicates the element-wise square $g_t^2 \odot g_t$ or cube $g_t \odot g_t \odot g_t$. Set $\alpha = [1e-2, 2e-5]$, $\beta_1 = 0.9, \beta_2 = 0.999, \beta_3 = 0.999987$, and $\epsilon = 10^{-8}$. All operations on vectors are element-wise. With β_1^t, β_2^t and β_3^t we denote β_1 β_2 and β_3 to the power t.

Require:α : Stepsize

Require: $\beta_1, \beta_2, \beta_3 \in [0, 1]$: Exponential decay rates for the moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0 : Initial parameter vector

$m_0 \leftarrow 0$ (Initialize $1^{\textbf{st}}$ moment vector)

$v_0 \leftarrow 0$ (Initialize 2^{nd} moment vector)

$u_0 \leftarrow 0$ (Initialize 3^{rd} moment vector)

$t \leftarrow 0$ (Initialize timestep)

while θ_t not converged do

$t \leftarrow t + 1$

$g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep t)

$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)

$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate)

$u_t \leftarrow \beta_3 \cdot u_{t-1} + (1 - \beta_3) \cdot g_t^3$ (Update biased third raw moment estimate)

$\widehat{m}_t \leftarrow m_t / \left(1 - \beta_1^t\right)$ (Compute bias-corrected first moment estimate)

$\widehat{v}_t \leftarrow v_t / \left(1 - \beta_2^t\right)$ (Compute bias-corrected second raw moment estimate)

$\widehat{u}_t \leftarrow u_t / \left(1 - \beta_3^t\right)$ (Compute bias-corrected third raw moment estimate)

$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \widehat{m}_t / \left(\sqrt{\widehat{v}_t} + (\sqrt[3]{\widehat{u}_t} * \epsilon)\right)$ (Update parameters)

end while

return θ_t(Resulting parameters)