CS795/895: Fundamentals of Deep Learning (Spring 2024)

Homework Assignment 4

Name: Sushmitha Halli Sudhakara

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 built-in 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 3, select the best model (with and without regularization) for each dataset. You should have a total of 8 models for one dataset you selected (one dataset, four optimizers, and with/without regularization).

Algorithm 1 Algorithm: Stochastic optimization. here g_t^2 denotes the element-wise square $g_t \odot g_t$

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 1st moment vector)
  $v_0 \leftarrow 0$ (Initialize 2nd moment vector)
  $u_0 \leftarrow 0$ (Initialize 2nd moment vector)
  $t \leftarrow 0$ (Initialize timestep)
  while θ_t not convereged **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)
    $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)
    $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)
    $\hat{u}_t \leftarrow u_t / (1 - \beta_3^t)$ (Compute bias-corrected third raw moment)
    $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / \left(\sqrt{\hat{v}_t} + (\sqrt[3]{\hat{u}_t} \cdot \epsilon)\right)$
  end while
  return θ_t (Resulting parameters)

Solution

Colab Notebook Link