February-June 2023 Semester
CS671: Deep Learning and Application
Programming Assignment 3

Date: March 09, 2023

Deadline for submission of code and report: **Friday, March 17, 2022, 10:00 PM**

The objective of this programming assignment is to deepen your understanding of the **optimizers for backpropagation algorithms**. Task is to train the fully connected neural network (FCNN) using different optimizers for the backpropagation algorithm and compare the number of epochs that it takes for convergence along with their classification performance.

You are given the subset of the MNIST digit dataset for the same. Each group is given 5 classes. Given dataset is train-validation-test separated. Every image is of the size 28 x 28. **Flatten each image to represent it as a vector of 784-dimension (28 x 28)**. The tasks for this assignment are as follows:

Develop an FCNN with different hidden layers (minimum number of hidden layers is 3 and maximum is 5).

a) Use cross-entropy loss.
b) Experiment with different a number of nodes in each of the layers.
c) Train each of the architectures using:
   i. Stochastic gradient descent (SGD) algorithm - (`batch_size`=1),
   ii. Batch gradient descent algorithm (vanilla gradient descent) – (`batch_size`=total number of training examples),
   iii. SGD with momentum (generalized delta rule) – (`batch_size`=1),
   iv. SGD with momentum (NAG) – (`batch_size`=1),
   v. AdaGrad – (`batch_size`=1),
   vi. RMSProp – (`batch_size`=1), and
   vii. Adam optimizer – (`batch_size`=1).
d) Use the **same initial random values of weights** for each architecture using each of the optimizers.
e) Use the ***absolute difference between average error of successive epochs fall below a threshold 10^{-4}*** as stopping criteria. **Do not use number of iterations as stopping criteria**.
f) Consider the learning rate (η) as 0.001 for all the optimizers.
g) Consider momentum parameter as 0.9 for both generalized delta and NAG.
h) Consider $\beta = 0.99$ and $\varepsilon = 10^{-8}$ for RMSProp.
i) Consider $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\varepsilon = 10^{-8}$ for Adam optimizer.

Presentation of Results:

1. Observe the number of epochs considered for convergence for each of the architectures. Tabulate and compare the number of epochs considered by each of the optimizers for each architecture.

2. Present the plots of average training error (y-axis) vs. epochs (x-axis) for each architecture. Superimpose the plots from each of the optimizers.
3. Give the training accuracy and validation accuracy for each of the optimizers in each of the architectures.
4. Choose the best architecture based on validation accuracy. Give the test confusion matrix and test classification accuracy along with training accuracy and confusion matrix for the chosen best architecture.

Each group of students must use the dataset identified for that group only.

You can use deep learning APIs (Tensorflow, PyTorch, Keras, etc.).

Report should be in PDF form and report by a team should also include the observations about the results of studies.

Instruction:

Upload in Moodle all your codes in a single zip file. Note that code(s) should be in a .py file, if you are coding in Python.

- **Give the name of the code folder as Group<*number*>_Assignment3_code**
  Example: Group01_Assignment3_code.
- **Give the name of the zip file as Group<*number*>_Assignment3_code.zip**
  Example: Group01_Assignment3_code.zip

Upload the report as PDF file.

- **Give the name to the report file as Group<*number*>_Assignment3_report.pdf**
  Example: Group01_Assignment3_report.pdf

We will not accept the submission if you don't follow the above instructions.