Discussion on the Project of Comparing Various Stochastic Gradient Methods

Part 1 I

- For this part you should get exactly the same result
- With same initial weights and the same operations you should get the same weights in the first several iterations
- For example, here are the weights of running `mnist` with the following parameters (scripts and results were generated by our TAs)

```
python3 script.py --optim SGD --bsize 256 --
--seed 42 --net CNN_4layers --train_set
/tmp3/data/mnist.mat --val_set /tmp3/data/mr
--dim 28 28 1
```

Part 1 II

- For simpleNN, first layer of running 11 batches are

  ```
  batch 1: 0.14049198 -0.03910705  0.18319398
  ...
  batch 11: 1.36893839e-01 -2.44279262e-02  1.
  ```

  Results of using Tensorflow

  ```
  batch 1: 0.14049198 -0.03910705  0.18319398
  ...
  batch 11: 1.36893839e-01 -2.44279262e-02  1.
  ```

Part 2 I

- You must think about how to clearly present your results

- For example, a table may be better than the following description:

  learning rate ?? gives final accuracy ??, best accuracy ??, learning rate ?? gives final accuracy ??, best accuracy ??, learning rate ?? gives final accuracy ??, best accuracy ??,

- You can see that "learning rate," "final accuracy," etc. appear many times

Part 2 II

- If the method fails to converge and get bad accuracy, from our discussion, you should decrease the learning rate

- For example, some tried Adam with learning rates 0.01, 0.1, 0.5 on cifar10, and all failed

- In this situation you should try for example 0.005 or 0.001

- Some reported higher accuracy than others. Then we realize that they used the same data set for training and prediction. This gives the training but not test accuracy

The Nan Issue in Project 1 I

- In project 1, some observed that under certain seed and learning rate, "nan" (not a number) occurred
- TAs printed out weights and found that after a few iterations, weights become very large
- Eventually things explode
- This is indeed a situation where SG diverges
- If the learning rate is decreased, then this issue should disappear

Other Comments I

- Please have both ID and name on the first page
- You should use the latex template
- I have no choice but to give you 0 point as we have said that twice (FAQ and project 1 discussion)
- For reports exceeding the page limit, this time 10 points are taken out
- Next time you will get 0 directly
- The quality of your reports significantly varies. Some are excellent but some are poor.
- Please do make sure that you pay enough efforts on taking this course