Discussion on the Project of Comparing Various Stochastic Gradient Methods by PyTorch
Many of you wrote a good report about your experiments and your findings. This is great.

We have the following test accuracy values for Adam with normalized data:

\[
\leq 50\%, 73\%, 76\%, 82\%
\]

It’s unclear what happened. I will get the code from some of you and conduct a study.

For running time, the main cost of each epoch is on calculating the sub-gradient.

So the time per epoch of these methods should be similar.

That’s the observation of most people.
But some have

SGD: 4,372 seconds
Adam: 3,833 seconds

This is a bit strange

I would like to stress again that your report should be no more than 2 pages

Don’t put any appendix.

Please use our format. You want to note that in any conference submission, papers not in their format are immediately rejected

Please spend time on thinking about what to put in your report
For example, will a table or a figure be better to show the result?

Note that organization is very important in any paper/report writing