Project: Comparing Various Stochastic Gradient Methods by PyTorch
Goal

- We check the use of several stochastic gradient implementations in PyTorch for training CNN
In the previous project, we tried only PyTorch’s simple stochastic gradient implementation.

Now we have discussed other variants.

Let’s try them in this project:

- Simple stochastic gradient (your previous project)
- Stochastic gradient with momentum
- Adagrad
- Adam

All settings (e.g., architectures, activation functions, etc.) are the same as the previous project.
Use their default parameter settings and initial solutions (so these methods may use different initial solutions)

Let’s run 50 epochs and check the relationship between accuracy and the accumulated # of epochs

A purpose is to see if in general a validation procedure is needed to decide the stopping condition

No tuning on other parameters yet

In your report, give your observations and thoughts