Teachers: J. Hennebert, M. Melchior
C. Gisler, P. Matvulj, O. Zayene

Practical Work 05 – 21/03/2024
Optimisers, Benefits of Depth

Objectives

The main objective of this PW is to explore the behavior of the different optimisers for training and to implement a learning rate schedule. Learn also how to set up and use "Weights and Biases" or "Tensorboard".

Submission

— **Deadline** : Wednesdays 3th April, noon

— **Format** : ipynb

Exercise 1 Characteristics of Optimisers with 2d Example

Analyse the optimisation paths of the optimisers discussed in the lecture by using a 2d example, i.e. SGD, Momentum, Momentum with Nesterov, RMSprop, Adam.

Use the utilities and instructions in the notebook `PW05_optimisers_2d_example_stud.ipynb`.

Exercise 2 Integrate with Monitoring Tool

Select one of the proposed monitoring tools, i.e. Weights & Biases or Tensorboard and bring yourself up to speed in using them. There are many different resources available on the web that instruct you how to integrate with them, e.g. Weights & Biases or Tensorboard.

You can use the selected tool for getting an overview on the results obtained in the next exercises. Make sure, however, that you provide the results as requested in the exercises in the notebooks.

Exercise 3 Prepare and Test Baseline

Set up the baseline for classifying Fashion MNIST images by using two CNN layers with max pooling, followed by one Dense layer and an output classification layer. Follow the "Steps to Test and Tune a Model" presented in the lecture. Use SGD without momentum and batchsize 64. Tune the learning rate.

Use the notebook `PW05_optimisers_FashionMNIST_stud.ipynb` as skeleton.

Exercise 4 Cross Validation

Use 5-fold cross-validation to estimate the validation accuracy and its error bars when training the FashionMNIST classifier of the previous exercise.

Again use the notebook `PW05_optimisers_FashionMNIST_stud.ipynb` as skeleton.

Exercise 5 Comparison of Different Optimisers for Training a FashionMNIST Classifier

Train the same CNN model as in the previous two exercises, but now by using different optimisers as learned during the lecture (and considered in the first exercise) : SGD, Momentum, Momentum with Nesterov, RMSprop, Adam.

Again use the notebook `PW05_optimisers_FashionMNIST_stud.ipynb` as skeleton.

Exercise 6 Learning Rate Schedule

For training again the same model for FashionMNIST classification implement a learning rate schedule.

Again use the notebook `PW05_optimisers_FashionMNIST_stud.ipynb` as skeleton.

Exercise 7 Review Questions (Optional)

— Summarise the steps to test and tune a model as presented in the lecture.

— Explain the mechanics of n-fold cross validation and in what sense it exploits well the available training / validation data.

— Explain why it is possible to obtain an estimator for the error bar for the validation accuracy, but not for the training accuracy.

— Summarise the claim that the 'Universal Approximation Theorem' states.

— Explain the term 'curse of dimensionality'.

— Provide an intuition on why deep models are better suited for representing the distribution of complex, unstructured data.

— Explain in words the difference between Momentum and RMSProp. How do you expect to choose the learning rate with Momentum and with RMSProp as compared with SGD ? Explain !

— Why is the learning used for Adam considered more independent of the particular model to be trained as compared with SGD ?