Computer Lab: Digit recognition
SD-TSIA 211

Tamim El Ahmad, Olivier Fercoq, Joël Garde, Iyad Walwil
27 January 2023

1 Submission and grading information

You can do the computer lab alone or in pairs. Please write a report and post it on `e-campus`. You can do it as a jupyter notebook or a pdf file.

Then, each of you will have to evaluate a couple of other students' reports and give comments.

Only the fact that you produce a report and evaluate your peers count in the final grade, so do not worry if you do not finish everything.
- 1 point for being present on the day of the lab
- 1 point for submitting a report
- 1 point for commenting 2 reports

2 Dataset

We are going to use the MNIST dataset, which consists of images of written digits together with their label. As it is a very common dataset, it can be downloaded directly with a function of the Keras library.

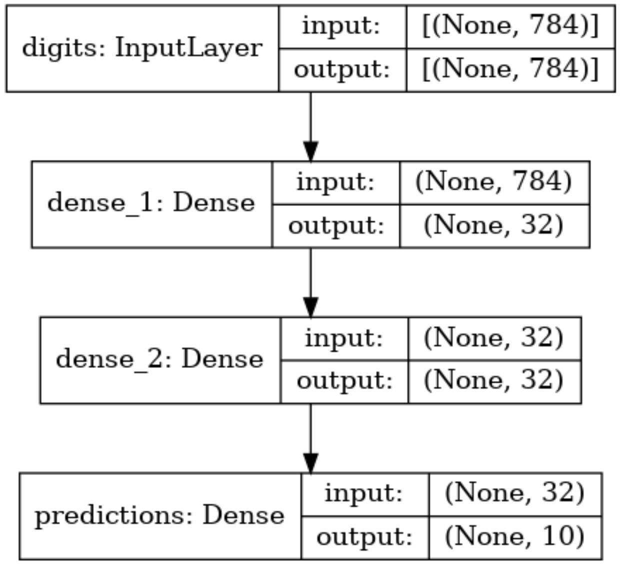
3 Dependences

We will use the following packages: scipy, numpy, matplotlib, tensorflow and keras. You will need them to be installed in order to run the code.

On Télécom Paris's computers, you should have all the packages available by using `/cal/softs/anaconda/anaconda3/bin/jupyter notebook`

4 Model

The file `TP_MNIST_basic_functions.ipynb` or `TP_MNIST_basic_functions.py` contains a neural network model written with Keras for the question of hand-written digit recognition.



The loss function is the categorical cross-entropy function whose formula is:

$$\ell(y^{\text{pred}}, y^{\text{true}}) = -\sum_{i=1}^{n} \sum_{d=0}^{9} y_{i,d}^{\text{true}} \log(y_{i,d}^{\text{pred}})$$

where $y_{i,d}^{\text{true}} = 1$ if image i represents digit d and $y_{i,d}^{\text{true}} = 0$ otherwise.

Question 4.1
How many optimization variables are we going to train using this model?

5 Stochastic gradient descent

Question 5.1
Reusing the code provided in the basic functions file, implement stochastic gradient descent for the resolution of the neural network model.

Question 5.2
Run it for one pass over the data (also called one epoch) and with a learning rate that satisfies the conditions required by theory.

Plot the objective value as a function of the iterations.

Hint: Do not compute the objective value at each iteration, only from time to time. This is sufficient for plotting purposes and will save computing time.

6 Empirical risk minimization

When we want to put more energy on the database we have, we can run the algorithm for more than one epoch. In this case, we consider that we are solving the training problem on a finite sample and we try to minimize the empirical risk.

Question 6.1
Implement stochastic gradient descent for the resolution of the empirical risk minimization problem and run it for 10 epochs.

Sometimes practicioners select several samples at each iteration (this is called a batch of samples) and they average them. This results in the algorithm

$$\begin{aligned}&\text{Select a batch } b_{k+1} \text{ of samples}\\ &x_{k+1} = x_k - \gamma_k \frac{1}{|b_{k+1}|} \sum_{i \in b_{k+1}} \nabla f(x_k, i)\end{aligned}$$

Question 6.2
According to you, what is the advantage of such a scheme?

Question 6.3
Add the option of the batch size to your algorithm.

7 Evaluation of the model

Question 7.1
Evaluate the accuracy of the model on the training set and on the test set.