Assignment 1

Advanced Deep Learning
27 September 2022

Exercise 1. *Follow the instructions in the python code* `mnist-assignment.py` *to complete a working program to train a MLP that solves the task of classification of handwritten characters.*

Exercise 2. *Once you have successfully completed Exercise 1 answer the following questions.*

Q1. On line 49 of the code it is commented that we want to perform a GD based optimization. However on line 45 we invoked `optim.SGD` as the optimizer. Explain why in this case we are still performing a gradient descent on the whole dataset even if it seems that we are invoking a stochastic method.

Q2. Discuss over-fitting issues by monitoring the train and test error curves.

Q3. Discuss what role does the choice of the network (i.e. number of layers and number of neurons per layer) have on the bias-variance trade-off. First describe your expectations based on theoretical analysis (arguing on the different capacity of the models) then test this expectations with a small experimental campaign. Is the expected behavior confirmed by experimental results? Briefly discuss your findings.

Q4. Discuss the benefits of using a cross entropy loss with respect to a quadratic loss.

Q5. Why using a one-hot encoding? Wouldn't be simpler to use a single output? *Hint:* The answer has to do with the interplay between the loss and the *sigmoidal* activation functions.

Q6. Test the sensibility of the gradient descent method with respect to the learning rate.

Q7. With the network architecture that is described in Q3. of Exercise 1 (line 32–33 of the code) do you experience any problem related to the *vanishing of the gradient*? Why?

*   *   *

Instructions for the answers

1. You are expected to collect your answers in a report that should be delivered in the form of a single PDF file for each assignment.

2. You need to structure your report so that in any point it is clear what is the question you are answering to.

3. Clarity and usage of appropriate non-ambiguous notation will have an influence in the final grade.

4. The conciseness of the answers will be as well an important parameters that will be taken into account in the final grade. In particular you are supposed to give complete and satisfactory answers without going out of topic or arguing on facts or results that are not relevant. [The Merriam-Webster dictionary defines the adjective concise as *"free from all elaboration and superfluous detail"*.]

*   *   *

Useful Resources
Besides the slides of the lecture and the notes of the lecture here are some useful resources that I encourage you to use:

- Chapter 1 of our companion book to the ML texbook *Machine Learning, a Constrained Based Approach* 2nd edition that you can find here `https://sailab.diism.unisi.it/mlbook/` by clicking on "Get Companion Book". In this chapter you can find a very broad introduction to Machine Learning and some information to set up an environment to manage Python installations and get acquainted with PyTorch. This is still a draft so please do not circulate it.

- Some snippets of code to gain some familiarity with Python and PyThorch can be found in the same webpage `https://sailab.diism.unisi.it/mlbook/` in the section "Code"–"Chapter 1".
- Useful resources for the Python website `https://www.python.org` and the PyTorch website `https://pytorch.org` where you can find information on the installation and a documentation of the languages.

References

[1] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).

[2] V. N. Vapnik, *Statistical Learning Theory* (John Wiley & Sons, 1998).

[3] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).

[4] L. Ambrosio, N. Gigli, G. Savaré, Gradient Flows in Metric Spaces and in Spaces of Probability Measures (Birkhäusen Verlag, 2005) This is a very advanced and technical book, be careful!