Universität des Saarlandes
FR 4.7 Sprach- und Signalverarbeitung

Prof. Dietrich Klakow
Pattern and Speech Recognition

Exercise Sheet 7

Optimization for Training Deep Models

Deadline: 10.01.2017, 23:59

First Order Methods on MNIST

Exercise 7.1

In this exercise, you will use different first order optimization methods on MNIST dataset.

a) Download the code *multilayer_perceptron.py* from course website.

b) Run the code using following optimization schemes for 5 epochs: (i) gradient descent, (ii) gradient descent with momentum with momentum parameter = 0.5, (iii) AdaGrad with initial accumulator value = 0.1, (iv) RMSProp with decay = 0.9, mometum = 0. Submit the accuracy on test data for the four cases. *(4 pts)*

c) Why does gradient descent perform better than AdaGrad if the magnitude of gradients is large? Explain using the update rule equations for both the methods. *(1 pt)*

Visualization of Optimization Algorithms

Exercise 7.2

This task will help you visualize the behavior of different optmization techniques on the three dimensional error surface $f(x, y) = 3x^2 - y^2$

a) Implement gradient descent(gd) algorithm (or reuse the code from Assignment-2) with learning rate = 0.01 and $(x, y) = (5, -1)$ as the starting point. Run the code for 30 iterations and store the intermediate points. *(2 pts)*

b) Implement gd with momentum with $\alpha = 0.7$ while keeping other parameters same as above. Run the code for 30 iteratons and store the intermediate points. *(2 pts)*

c) Plot the 3D/contour plot of the function with paths of gd and gd with momentum using different colors. *(2 pts)*
For plotting in matlab, functions like *scatter3* can be used for 3D plot and *contour* for contour plot. In python, there are is a list of functions available, refer the link: http://matplotlib.org/mpl_toolkits/mplot3d/tutorial.html

d) Evaluate the function value at the termination points (after 30 iterations) for both the methods and produce the values. Which algorithm has made better progress in minimizing $f(x, y)$? *(1 pt)*

e) Implement Newton's method (or reuse the code from Assignment-2), run it for 5 iterations with same starting point and report your observations. Verify your observations by analytically computing Newton steps. This should explain why plain vanilla Newton's method is not suited for training deep neural networks. *(3 pts)*

Exercise 7.3

To understand the effectiveness of AdaGrad, consider $f(x, y) = 0.001x^2 - 0.001y^2$ which is a very flat function.

a) With learning rate = 0.1, and $(x, y) = (3, -1)$ as the starting point, run the *gd* algorithm for 300 iterations. *(1 pt)*

b) Implement the AdaGrad algorithm as described in the slide 21 of chapter-7 and run it with the same settings as above. Let $\delta = 10^{-9}$. *(3 pts)*

c) Plot the 3D/contour plot of the function with paths of *gd* and AdaGrad using different colors. *(1 pts)*

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

Submission architecture

You have to generate a **single ZIP file** respecting the following architecture:

```
tutorial2_<matriculation_nb1>_<matriculation_nb2>_<matriculation_nb3>
|
+--- source
|      |
|      +------- file 1
|      +------- file 2
|      +------- ...
+--- report.pdf
+--- README.txt
```

where

- **source** contains the source code of your project,
- **report.pdf** is the report where you present your solution with **the explanations** and the plots.

- **README** which contains group member informations (name, matriculation numbers and emails) and a **clear** explanation about how to compile and run your source code

The ZIP filename has to be :

`tutorial7_<matriculation_nb1>_<matriculation_nb2>_<matriculation_nb3>.zip`

Some hints

We advice you to follow the following guidelines in order to avoid problems :

- Avoid building complex systems. The exercises are simple enough.
- Do not include any executables in your submission, as this will cause the e-mail server to reject it.

Grading

Send your assignment to the tutor who is responsible of your group:

- Merlin Köhler s9mnkoeh@stud.uni-saarland.de
- Goutam Y G goutamyg@lsv.uni-saarland.de
- Ahmad Taie s8ahtaie@stud.uni-saarland.de

The email subject should start with `[PSR TUTORIAL 7]`