POLITECNICO DI TORINO
NUMERICAL OPTIMIZATION FOR LARGE SCALE PROBLEMS AND STOCHASTIC OPTIMIZATION

Problem list, stochastic optimization
a.y. 2019/20

Please choose one of the following problems. Whatever problem you choose, I expect you to submit a single pdf file, splitted into two parts. A first part should contain a commented script in your favorite programming language that implements an algorithm and test it on a simple problem (see the problem list below). Please make sure to use sensible names for the variables and functions, and to provide enough comments and explanations to render the code readable to a non expert of the specific language.

The second part should contain a summary of the results you got from running your script. Please explain and comment if you are satisfied with the results you got or if it seems that something went wrong. Be aware that I may not be able to execute the code, therefore include all the numbers and figures that you obtain as an output, before commenting on them.

Please make every effort to make the whole submission understandable.

1. Neural network, Section 4.2 Gosavi's textbook

Please replicate Example D, page 64, between Section 4.2.7 and Section 4.2.8, in the Second Edition of the text book by Abhijit Gosavi. In order to avoid the explosion of the weights, instead of updating the learning rate with the rules explained in the textbook, start with a small value, e.g. $\mu = 0.01$, and at each iteration, update μ according to the following rule

$$\mu \leftarrow \mu \cdot 0.99999$$

or just keep it constant to $\mu = 0.01$. In order to fix a stopping rule you can either fix a tolerance or just run the algorithm for a fixed number of iterations (500 is fine). You may play with the other tuning parameters if you wish.

2. Stochastic steepest descent

Write a function of two arguments (x, y) that outputs the value of the polynomial $f(x, y) = 18 + 11.4x - 31x^2 + 0.6x^3 + x^4 + 50y^2 = (x - 5)(x-1)(x+0.6)(x+6)+50y^2$ perturbed with a random Gaussian noise with zero mean and standard deviation equal to 15. In order to find the minimum of this function, apply to it a stochastic gradient descent method. To estimate the gradient, use the method of simultaneous perturbations. Use several random initial values in the range $x \in [-6, 5]$, $y \in [-5, 5]$. Discuss if you have been able to find the desired minimum.

3. Nelder-Mead algorithm

Write a function of two arguments (x, y) that outputs the value of the polynomial $f(x, y) = 18 + 11.4x - 31x^2 + 0.6x^3 + x^4 + 50y^2 = (x-5)(x-1)(x+0.6)(x+6)+50y^2$ perturbed with a random Gaussian noise with zero mean and standard deviation equal to 15. In order to find the minimum of this function, apply to it the Nelder-Mead algorithm. Use several random initial values for the vertices of the initial triangle, in the range $x \in [-6, 5]$, $y\epsilon[-5, 5]$. Discuss if you have been able to find the desired minimum.

4. Reinforcement learning

Let us consider the following system. A small shop has two employees. One is always serving customers, the other is doing other jobs as well. Suppose that the number of customers in the shop is between 0 and 6, since any further incoming customer would not enter the shop if 6 people are already in. Action 1 correspond to let the second employee do his other jobs and having a single employee serving the customers. Under this action the system follows the following dynamics. At each unit of time if the current number of customers is between one and 5, there is equal probability that

- a new customer enter the shop
- nothing happens
- a customer is served and exit the shop

If there is no customer in the shop, with probability $2/3$ nothing happens in the next unit of time and with probability $1/3$ a new customer enters the shop. If there are 6 customers in the shop, with probability $2/3$ nothing happens in the next unit of time and with probability $1/3$ a customer will be served and exit the shop. Rewards are discounted by a factor $\lambda = 0.9$. There is an immediate reward of 1 unit for each customer served (whatever the number of customers in queue). If the states are numbered from 1 to 7, and at state i we have $i - 1$ customers in queue, the reward matrix is

$$R1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Action 2 is instead to have both employees serving the customers. in this case the dynamics would be changed as follows. At each unit of time if the current number of customers is between one and 5,

- a new customer enter the shop with probability $1/4$
- nothing happens with probability $1/4$
- a customer is served and exit the shop with probability $1/2$

If there is no customer in the shop, with probability $2/3$ nothing happens in the next unit of time and with probability $1/3$ a new customer enters the shop. If there are 6 customers in the shop, with probability $1/3$ nothing happens in the next unit of time and with probability $2/3$ a customer will be served and exit the shop. Whit respect to the previous reward structure, under this action there is an additional fixed cost per unit of time of 0.2 units due to the fact that, while serving, the second employee is not doing the other jobs. Therefore the reward matrix is

$$R2 = \begin{pmatrix} -0.2 & -0.2 & -0.2 & -0.2 & -0.2 & -0.2 & -0.2 \\ 0.8 & -0.2 & -0.2 & -0.2 & -0.2 & -0.2 & -0.2 \\ -0.2 & 0.8 & -0.2 & -0.2 & -0.2 & -0.2 & -0.2 \\ -0.2 & -0.2 & 0.8 & -0.2 & -0.2 & -0.2 & -0.2 \\ -0.2 & -0.2 & -0.2 & 0.8 & -0.2 & -0.2 & -0.2 \\ -0.2 & -0.2 & -0.2 & -0.2 & 0.8 & -0.2 & -0.2 \\ -0.2 & -0.2 & -0.2 & -0.2 & -0.2 & 0.8 & -0.2 \end{pmatrix}$$

Please use Reinforcement Learning to learn the best policy. As a step-size rule, please use

$$\alpha_k = \frac{A}{B+k}$$

with $A = 150$ and $B = 300$. The number of iterations can be 10^5. Use as initial state the one where no customer is in the shop.