Deadline: Sun Nov 28, 2021, 8:00 am Submit single unzipped PDF file on learn-web course "SoSe 2021: 3104 Modern Optimization Techniques

Instructions

Please following these instructions for solving and submitting the exercise sheet.

1. Student should clearly write his/her name, matriculation number and tutorial group number (i.e. "Group 1: Tuesday Tutorial", "Group 2: Wednesday Tutorial").
2. The submission should be made before the deadline, only through learnweb to your group submission link.
3. Should be submitted as a single unzipped PDF file on learn-web course "SoSe 2021: 3104 Modern Optimization Techniques".
4. Each student must submit an individual solution in-order to be eligible for bonus points.
5. Group submission are acceptable but will not contribute towards bonus points.

1 Linear Regression with Gradient Descent (10 points)

Suppose that there is a research group wants to measure the frequency and effects of shouting and cursing while driving a car. They gave you some data and you have the intuition that it behaves linearly.

Shouting Frequency	Cursing Frequency	Traffic Accidents Frequency (Response)
1.5	2	10
3	2.5	15.5
4.5	3	21

Therefore, we need to find the parameter vector β that minimizes the loss over all instances x_i:

$$\mathcal{L}(X, \beta, y) = \sum_{i=1}^{3} (\beta^\top x_i - y_i)^2$$

a) Is it possible to obtain closed form solution? If yes, find it? [3 pt]

b) Explain in your own words, why we apply an approximate learning algorithm to a problem where an analytical solution exists? [2 pt]

c) Assume your model is initialized by $\beta = (1, 1, 1)^T$, perform 2 iterations of GD to compute the updates of β with a step size of $\mu = 0.1$. What are the errors and the overall loss after updating β? [5 pt]

2 Linear Regression with Stochastic Gradient Descent & Adagrad (10 points)

For this question we will re-use the problem settings X and y define in Question 1. Answer the following questions:

a) Explain in your own words, what is the difference of stochastic gradient descent compared to a normal gradient descent? [2 pt]

b) Do two epochs using stochastic gradient descent with a step size of $\mu = 0.1$ and report the errors and total loss after each epoch, with an initial $\beta = (1, 1, 1)$. **Please go over the instances in order, i.e. first line, second line, third line of X.** [4 pt]

c) Repeat the same procedure by using a stochastic gradient descent with Adagrad for an initial step size of $\mu = 0.1$. Does Adagrad help? [4 pt]