Modern Optimization Techniques

First Take-home Exam

Prof. Dr. Dr. Lars Schmidt-Thieme
M.Sc. Eya Boumaiza

Nov 20^{st}, 2020
ISMLL Universität Hildesheim

Note:

- Time: 240 minutes
- Add your name and matriculation number on top of every page.
- Please provide clear and detailed answers to get full points.
- Please make sure the provided solution is clearly written and scanned.
- Solutions need to be submitted before the deadline. Please consider submitting their solutions 5 minutes before the deadline in order to ensure that the solution is uploaded and no internet problem interrupts it.
- Plagiarism, cheating and group submissions are not allowed. Any suspicious solution will be further investigated and the student will fail the course for this semester in case we prove it.

1. Convexity and Simplex

1A. Convexity of Sets (3 points)

Show, using the definition, the convexity of the set given below. Discuss if it is an affine set using the definition as well. What is the shape of this set (How it looks like)?

$$S = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 \leq r^2\}$$

1B. Simplex Method (4 points)

The owner of a shop producing electronics wishes to determine the best mix for his four products: Smartphones, Laptops, Tablets and TVs. His shop is limited to working 4 days/week on assembling GPU cards, 3 days/week on connecting the screens of the products and 3 days/week on setting up the battery for these products. The following table indicates production data for the electronics.

Usage per unit of trailer					Resources
	Smartphone	Laptop	TV	Tablet	availabilities
GPU working days	1	3	0	1	4
Screen working days	0	1	4	1	3
Battery working days	2	1	0	1	3
Profit (×1000 €)	2	4	1	1	

The owner wants to maximize the production of the electronics in his shop. **Your tasks are:**

a. Define the objective function and the associated constraints.

b. Find the optimal solution for the problem setting using the simplex method (with detailed explanation of each step).

1C. Convexity of Functions (3 points)

a. Discuss the convexity of the function h given below on its domain. i.e., is it convex, strictly convex, strongly convex ? Does the function has a minimum on this domain? why?

$$h : (0, +\infty) \longrightarrow \mathbb{R}$$

$$h(x) = x^2$$

b. Suppose $f : \mathbb{R} \longrightarrow \mathbb{R}$ is an increasing function that is convex but does not have a minimum on its domain (a, b). Let g denote its inverse, i.e., the function with domain $(f(a), f(b))$ and $g(f(x)) = x$ for $a < x < b$. What can you say about convexity or concavity of g? (provide proof with all possible cases)

2. Unconstrained Optimization: Gradient Descent

2A. Closed form solution and approximative learning (2 points)

Derive the closed form solution of a ridge regression. Why an approximative learning algorithm like GD is applied to problems where an analytical solution exists?

2B. Gradient Descent Method (4 points)

Suppose you are a Data Scientist in a group who wants to win the Ig Nobel Prize. For this purpose, your research group wants to predict the change in the Blood Oxygen Level of a patient after two days of being diagnosed with a Corona Virus based on the change after the first day in the blood pressure and the temperature. They gave you some data and you have the intuition that it behaves linearly.

Change in Temperature	Change in Blood Pressure	Change in Blood Oxygen Level
2	0	-1
3	1	5
-1	2	7

You will use GD method to learn a ridge regression model using the objective function given below **(don't forget to add the bias)**.

$$\mathcal{L}_{ridge}(X, Y, \beta) = \|X\beta - Y\|_2^2 + \lambda\|\beta\|_2$$

Your tasks are:

a. Derive the update rule for a β^{t+1}.

b. Perform 2 iterations of the GD with an initial $\beta^0 = (-1, 1, 2)^T$, a step size $\mu = 0.1$ and a regularization weight $\lambda = 0.1$.

c. At the end of each iteration, calculate the value of the loss function. Describe what happens with the error.

2C. Gradient Descent Method with Backtracking Line Search (4 points)

Since the choice of the step size μ is crucial when using the GD method, you will use Backtracking Line Search as a step size method for the GD method. The general Backtracking Condition is as follows:

$$f(x + \mu\Delta x) > f(x) + a\mu\nabla f(x)\Delta x$$

Your tasks are:

a. Derive the Backtracking Condition for the ridge loss function given above using the negative gradient $\Delta x = -\nabla f(x)$ as descent direction.

b. Perform 2 iterations of the GD with Backtracking Line Search with an initial $\beta^0 = (-1, 1, 2)^T$, a regularization weight $\lambda = 0.1$ and $a = b = 0.1$

c. At the end of each iteration, calculate the value of the loss function. Describe what happens with the error. Does the Backtracking Line Search method help?

3. Unconstrained Optimization: Stochastic Gradient Descent Method

3A. Stochastic Gradient Descent vs Gradient Descent (3 points)

a. Explain in your own words, what is the difference of stochastic gradient descent as compared to a normal gradient descent.

b. What are the changes made on the Backtracking Line Search method when used with SGD method? Describe 2 alternative methods to the Backtracking Line Search method when used with SGD and explain how they differ from the Backtracking Line Search.

3B. Stochastic Gradient Descent Method (4 points)

Using the same data from the previous problem, you will use SGD method, instead of GD, to learn a ridge regression model using the same objective function given above **(don't forget to add the bias)**.

Your tasks are:

a. Derive the update rule for a β^{t+1}.

b. Perform 3 iterations of the SGD with an initial $\beta^0 = (-1, 1, 2)^T$, a step size $\mu = 0.1$ and a regularization weight $\lambda = 0.1$. **Please go over the instances in order, i.e. first line, second line, third line of X.**

c. At the end of each iteration, calculate the value of the loss function. Describe what happens with the error?

3C. Stochastic Gradient Descent Method with Adagrad (3 points)

Here, you will use Adagrad as a step size method for the GD method.

Your tasks are:

a. Perform 3 iterations of the SGD with Adagrad with an initial $\beta^0 = (-1, 1, 2)^T$, an intial step size $\mu_0 = 0.1$ anf a regularization weight $\lambda = 0.1$.

b. At the end of each iteration, calculate the value of the loss function. Describe what happens with the error. Does the Adagrad method help?