Modern Optimization Techniques - Exercise Sheet 2

Lydia Voß

voss@ismll.de

November 4, 2016

Solutions need to be handed in until **Thursday, November 10th, 2016 at 12:00**

Exercise 1: Linear Regression with Gradient Descent (13P)

We want to learn a non-regularized linear regression model using gradient descent on the data given by the design matrix A and labels y:

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad y = \begin{pmatrix} 11 \\ 10 \\ 8 \end{pmatrix}$$

Therefore, we need to find the parameter vector $\beta = (\beta_0, \beta_1, \beta_2)$ that minimizes the loss over all instances a_i:

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

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

b) Compute the closed form solution for a non-regularized linear regression optimized for least squares!

c) Assume your model is initialized by $\beta = (1, 1, 1)$, compute the errors

$$e_i = \beta^\top a_i - y_i$$

and the overall loss of the model.

d) Using the error terms, compute the updates of β with a step size of $\mu = 0.1$. What are the errors and the overall loss after updating once?

Exercise 2: Backtracking Line Search (7P)

Let us define a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ through:

$$f(x_1, x_2) = x_1^2 + x_2^2$$

a) Suppose you want to do a backtracking line search using the negative gradient $\Delta x = -\nabla f(x)$ as descent direction. Suppose you are in a current point $x' = (x_1', x_2')$, write down the backtracking condition

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

for these special settings.

b) We pick $a = 0.5$, $b = 0.1$ and start with a rather high initial step size $\mu = 10$. How small does μ have to become for the backtracking condition to be false? How many backtracking iterations will be done until this happens?