Modern Optimization Techniques - Exercise Sheet 4

Lydia Voß

November 20, 2017

Solutions need to be handed in until **Monday, November 27th, 2017 at 10:00** via the postboxes (preferred) otherwise Learnweb

Exercise 1: Linear Regression with Stochastic Gradient Descent & Adagrad (12P)

Let us revisit our toy linear regression example from last time with data given by design matrix A and labels y:

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

We want 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, what is the difference of stochastic gradient descent compared to a normal gradient descent!

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 A.**

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?

Exercise 2: Logistic Regression (8P)

The logistic regression learns a linear regression model $f(x) = \beta^\top x$ by optimizing the logistic loss function over data D:

$$\mathcal{L}(y, \beta, D) = \sum_{x \in D} \log(1 + \exp(-y \cdot \beta^\top x))$$

a) Compute the gradient of $\mathcal{L}(y, \beta, D)$ with respect to β!

b) Write down the pseudocode for a GD and an SGD that learns the logistic regression!

(By pseudocode it is meant that you do the derivation of the loss with respect to the model parameters β out of a) and then write down the update formula and so on for this special model, do not just copy the general formulas out of the lecture!!)