Modern Optimization Techniques - Exercise Sheet 7

Lydia Voß

voss@ismll.de

December 16, 2016

Solutions need to be handed in until **Thursday, January 12th, 2017 at 12:00**

Exercise 1: linear Regression with Coordinate Descent (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 Coordinate Descent compared to a normal Gradient Descent!

b) Do two epochs using coordinate descent. Report the errors and the overall loss after each epoch, with an initial $\beta = (1, 1, 1)$.

Exercise 2: Coordinate Descent (8P)

a) Show that coordinate descent fails for the function

$$g(x) = |x_1 - x_2| + 0.1(x_1 + x_2)$$

Hint: Verify that the algorithm terminates after one step while $\inf_x g(x) = -\infty$

b) Let

$$\mathcal{L}(x) = f(x) + \lambda \|x\|_1$$

be l_1-regularized minimization with $f(x)$ convex and differential and $\lambda \geq 0$. Assume we converge in a fixed point x^*. show that x^* is optimal, i.e. it minimizes $\mathcal{L}$.
Hint: Use the subdifferential you have seen in the previous lecture and exercise sheet. $\|x\|_1 = |x|$