Modern Optimization Techniques - Exercise Sheet 8

Lydia Voß

Solutions need to be handed in until **Monday, January 15th, 2018 at 10:00 am via Postbox or Learnweb**

Exercise 1: linear Regression with Coordinate Descent (10P)

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: Constrained Minimization (10P)

For the two following constrained problems, plot the level sets of f_0 and the given constraints to then graphically find $x^\star$.

a)

$$\begin{aligned} \text{minimize} \quad & f_0(x_1, x_2) = x_1^2 + x_2^2 \\ \text{subject to} \quad & h(x_1, x_2) = x_1 + 2x_2 = 3 \end{aligned}$$

Write down the KKT conditions for this optimization problem and analytically compute $x^\star$!

b)

$$\begin{aligned}
\text{minimize} \quad & f_0(x_1, x_2) = x_1 + x_2 \\
\text{subject to} \quad & h(x_1, x_2) = x_1 - x_2 = 2 \\
& f_1(x_1, x_2) = x_1 \geq 0 \\
& f_2(x_1, x_2) = x_2 \geq 0
\end{aligned}$$

Reason why you cannot compute the dual problem for a linear program as this one!

Bonus Exercises to earn extra points!

Bonus Exercise 1: Coordinate Descent (10P)

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. $\|x\|_1 = |x|$

Bonus Exercise 2: Newton Algorithm for Equality Constrained Problems (10P)

Let us again consider the following equality constrained optimization problem

$$\begin{array}{rl} \text{minimize} & f_0(x_1, x_2) = x_1^2 + x_2^2 \\ \text{subject to} & h(x_1, x_2) = x_1 + 2x_2 = 3 \end{array}$$

Optimize this problem using the Newton Algorithm for Equality Constrained Problems with a step size of $\mu = 1$. Start it once in the feasible point $x = (0, 1.5)$ and once in the non-feasible point $x = (0, -5)$. How many iterations does the algorithm need to converge? Explain your findings!