Optimization and Machine Learning

Midterm 2

December 8, 2010

- **Please give details of your calculation. A direct answer without explanation is not counted.**
- **Your answers must be in English.**
- **You can bring notes and the textbook. Other books or electronic devices are not allowed.**

Problem 1 (20%)

Consider the function

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

At the point $\boldsymbol{x}_k = [1, 0]^T$, find

(a) the gradient descent direction

(b) $\boldsymbol{x}_{k+1}$ by exact line search on the gradient descent direction

(c) the Newton direction

(d) $\boldsymbol{x}_{k+1}$ by exact line search on the Newton direction

Problem 2 (40%)

Consider the following quadratic function:

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T Q \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x}$$

Assume Q is symmetric and positive definite.

(a) What's the gradient of $f(\boldsymbol{x})$?

(b) Derive $\boldsymbol{x}_{k+1}$ by the gradient descent method with the exact line search. That is, represent $\boldsymbol{x}_{k+1}$ by a form of using $\boldsymbol{x}_k$

(c) Define

$$\|\boldsymbol{x}\|_Q^2 = \boldsymbol{x}^T Q \boldsymbol{x}.$$

Do we have

$$\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}^*\|_Q^2 = f(\boldsymbol{x}) - f(\boldsymbol{x}^*)?$$

We assume $\boldsymbol{x}$ is any vector and $\boldsymbol{x}^*$ is the minimum of $f(\boldsymbol{x})$. Prove or disprove the result.

(d) Represent $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|_Q^2 = C_k\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q^2$,

Where C_k is a scalar related to $\boldsymbol{x}_k$.

Hint: Prove and use the following property:

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\|_Q^2 = \nabla f(\boldsymbol{x}_k)^T Q^{-1} \nabla f(\boldsymbol{x}_k)$$

Problem 3 (40%)

Consider the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi},\rho,b} \quad & \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} - \nu\rho + \frac{1}{l}\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq \rho - \xi_i \\ & \xi_i \geq 0, \ \ \rho \geq 0 \end{aligned}$$

Note that $\nu \in [0, 1]$ is a positive constant.

The main difference from standard SVM is that a new non-negative variable ρ is introduced.

(a) Derive the dual problem.

(b) Show that the primal and dual problems have the same KKT conditions.

(c) If α_i is the Lagrangian multiplier of

$$y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq \rho - \xi_i,$$

Can you simplify the dual to have only one variable $\boldsymbol{\alpha}$?