Optimization and Machine Learning

Final Exam

January 7, 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 (50%)

Consider the following L1-regularized problem ($C > 0$)

$$\min_{\boldsymbol{w}} \quad \|\boldsymbol{w}\|_1 + \frac{C}{2}\sum_{i=1}^{l} \max(0, 1 - y_i \boldsymbol{w}^T \boldsymbol{x}_i)^2. \tag{1}$$

(a) Prove that it is equivalent to

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi}} \quad & \|\boldsymbol{w}\|_1 + \frac{C}{2}\sum_{i=1}^{l} \xi_i^2 \\ \text{s.t.} \quad & y_i \boldsymbol{w}^T \boldsymbol{x}_i \geq 1 - \xi_i \\ & \xi_i \geq 0. \end{aligned} \tag{2}$$

(b) Prove that $\xi_i \geq 0$ is not needed

(c) Rewrite (2) as

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\xi}} \quad & \boldsymbol{e}^T(\boldsymbol{w}^+ + \boldsymbol{w}^-) + \frac{C}{2}\sum_{i=1}^{l} \xi_i^2 \\ \text{s.t.} \quad & y_i(\boldsymbol{w}^+ - \boldsymbol{w}^-)^T \boldsymbol{x}_i \geq 1 - \xi_i \\ & w_i^+ \geq 0,\ w_i^- \geq 0. \end{aligned} \tag{3}$$

Derive the dual problem of (3). Simplify it to include a variable $\boldsymbol{\alpha}$, where α_i is the Lagrange multiplier of

$$y_i(\boldsymbol{w}^+ - \boldsymbol{w}^-)^T \boldsymbol{x}_i \geq 1 - \xi_i.$$

Problem 2 (50%)

Define the following operation:

$$(x-c)_+^2 = \begin{cases} 0 & \text{if } x \leq c \\ (x-c)^2 & \text{if } x \geq c \end{cases}$$

Consider the following function

$$\begin{aligned} f(x,y,z) = &- xy - yz - zx + (x-1)_+^2 + (-x-1)_+^2 \\ &+ (y-1)_+^2 + (-y-1)_+^2 + (z-1)_+^2 + (-z-1)_+^2 \end{aligned}$$

and we minimize it by a coordinate descent method using the order of $(x, y, z, x, y, z, \ldots)$

(a) If (y, z) is fixed, x is changed to which value?

Hint: the answer should be

$$(1 + \triangle \cdot \text{sign}(\square))\text{sign}(\square)$$

and you need to find $\triangle$ and $\square$, which are both expressions of y and z.

Please do this sub-problem carefully as you will need its answer for (b) and (c). You need to prove your result. A rough calculation does not count.

(b) If (x, z) is fixed, y is change to?

If (x, y) is fixed, z is change to?

(c) Starting from a point $(-1 - \xi, 1 + \frac{\xi}{2}, -1 - \frac{1}{4}\xi)$, run six steps of the coordinate descent method.

That is, each of x, y, and z is updated twice.

What is your observation?

Do you think that the coordinate descent method will converge to a single point?