Optimization, Written Assignment #5

May 14, 2008

1

Recall from recitation that if $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i)$, then its convex conjugate $f^*(y_1, \ldots, y_n) = \sum_{i=1}^{n} f_i^*(y_i)$. It is not, however, generally the case that if $f(\vec{x}) = \sum_{i=1}^{n} f_i(\vec{x})$, then $f^*(\vec{y}) = \sum_{i=1}^{n} f^*(\vec{y})$. Give an example in which $f(\vec{x}) = f_1(\vec{x}) + f_2(\vec{x})$, but $f^*(\vec{y}) \neq f_1^*(\vec{y}) + f_2^*(\vec{y})$.

2

In this problem we will consider a variant of logistic regression which imposes a bias term and weight regularization. Recall that the logistic loss is given by

$$g(x) = \log\left(1 + e^{-x}\right)$$

2.1

We first want to establish that

$$g^*(y) = \begin{cases} -h(-y) & \text{if } -1 \leq y \leq 0 \\ \infty & \text{if } y < -1 \text{ or } y > 0 \end{cases}$$

where $h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$, with $h(0) = h(1) = 0$ is the binary entropy function.
We saw in class that

$$g^*(y) = -h(-y)$$

for $-1 < y < 0$.

2.1.1

Show that $g^*(0) = g^*(-1) = 0$

2.1.2

Show that $g^*(y) = \infty$ for $y < -1$ or $y > 0$

2.2

We are now ready to consider the regularized logistic regression problem. We want to minimize both $\log_2(w, b) = \sum_{i=1}^{m} g(y_i(w' x_i + b))$ and $\|w\|_2$, where $w \in \mathbb{R}^n$, $b \in \mathbb{R}$ are our unconstrained optimization variables. We will consider various ways to solve this bicriterion problem.

2.2.1

Consider

$$\text{minimize} \quad : \quad \sum_i g\left(y_i\left(w^T x_i + b\right)\right) + \frac{\lambda}{2}\left\|w\right\|_2^2 \tag{1}$$

Where λ is a constant. Derive the dual of this problem.

2.2.2

Consider

$$\begin{aligned} \text{minimize} \quad &: \quad \sum_i g\left(y_i\left(w^T x_i + b\right)\right) \\ \text{subject to} \quad &: \quad \left\|w\right\|_2^2 \leq B \end{aligned} \tag{2}$$

Where B is a constant. Derive the dual of this problem.

2.2.3

Consider

$$\begin{aligned} \text{minimize} \quad &: \quad \left\|w\right\|_2^2 \\ \text{subject to} \quad &: \quad \sum_i g\left(y_i\left(w^T x_i + b\right)\right) \leq C \end{aligned} \tag{3}$$

Where C is a constant. Derive the dual of this problem.

2.2.4

Consider

$$\text{minimize} \quad : \quad \sum_i g\left(y_i\left(w^T x_i + b\right)\right) + r\left\|w\right\|_2 \tag{4}$$

Where r is a constant. Derive the dual of this problem.

2.3

We will consider two cases:

1. The optimal value of equation 1 with $\lambda = 0$ is 0
2. The optimal value of equation 1 with $\lambda = 0$ is positive

2.3.1

Prove that in both cases, the set of optimal solutions to equation 3 with $C \geq 0$, is exactly the set of Pareto-optimal solutions.

2.3.2

In what cases is the set of Pareto-optimal solutions exactly the set of optimal solutions of equation 2 with $B \geq 0$? Explain, and provide counterexamples.

2.3.3

In what cases is the set of Pareto-optimal solutions exactly the set of optimal solutions of equation 1 with $\lambda \geq 0$? Explain, and provide counterexamples.

2.3.4

In what cases is the set of Pareto-optimal solutions exactly the set of optimal solutions of equation 4 with $r \geq 0$? Explain, and provide counterexamples.