STATS 608A, Fall 2015
Homework 3

Ambuj Tewari

Assigned: Nov 25, 2015. Due: Dec 5, 2015.

1 Properties of self-concordant functions (5 points)

Recall the definition of a self-concordant function $f : \mathbb{R} \to \mathbb{R}$ from the class lecture. Prove the following:

1. If $f(x)$ is self-concordant then so is $f(ax + b)$ for $a \neq 0$.
2. If $f(x)$ is self-concordant then so is $af(x)$ for $a \geq 1$.
3. If f_1, f_2 are self-concordant then so is $f_1 + f_2$.
4. $f(x) = x \log x - \log x, x > 0$ is self-concordant.
5. $f(x) = e^x$ is not self-concordant.

2 Coordinate-wise minimization for the Lasso (5 points)

Consider the Lasso problem

$$\min_{\beta \in \mathbb{R}^p} f(\beta) := \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

Suppose the values of all coordinates of β, except β_i, are fixed at some values given by the $p-1$ dimensional vector β_{-i}. Also, let X_{-i} denotes the $n \times (p-1)$ submatrix of X obtained by removing the ith feature/covariate. In class, we saw that minimization of $f(\beta)$ over β_i with other coordinates fixed at β_i is achieved by setting

$$\beta_i = S_{\lambda/\|X_i\|_2^2}\left(\frac{X_i^\top (y - X_{-i}\beta_{-i})}{X_i^\top X_i}\right),$$

where $S_\tau(x) = \text{sign}(x)(|x| - \tau)_+$ is the soft thresholding function. Prove this claim rigorously.

3 Conditional gradient (Frank-Wolfe) convergence analysis (5 points)

When the conditional gradient method is used to solve the problem

$$\min_{x \in C} f(x) =: f^\star$$

we saw, in class, that the following key inequality holds:

$$f(x^{(k)}) \leq f(x^{(k-1)}) - \gamma_k g(x^{(k-1)}) + \frac{\gamma_k^2}{2} M$$

where $g(x) = \max_{s \in C} \nabla f(x)^\top (x - s)$ is an upper bound on the sub-optimality gap $f(x) - f^\star$ and M is the curvature constant defined as

$$M = \sup \left\{ \frac{2}{\gamma^2} \big(f(y) - f(x) - \nabla f(x)^\top (y - x) \big) \ : \ x, s, y \in C, y = (1-\gamma)x + \gamma s \right\}.$$

Setting $\gamma_k = 2/(k+1)$, show, by induction on k, that the key inequality implies

$$f(x^{(k)}) - f^\star \leq \frac{2M}{k+2} \ , \ k \geq 1.$$

4 Conditional gradient (Frank-Wolfe) for trace norm constrained problems (5 points)

Suppose we use conditional gradient method to solve

$$\min_{\|X\|_{\mathrm{tr}} \leq t} f(X)$$

where $\|X\|_{\mathrm{tr}}$ is the trace norm of a matrix $X \in \mathbb{R}^{m \times n}$ defines as the sum of the singular values $\sum_{i=1}^{\min\{m,n\}} \sigma_i(X)$. To compute the next iterate $X^{(k)}$ from the current iterate $X^{(k-1)}$, we will first need to compute

$$S^{(k-1)} = \operatorname*{argmin}_{\|S\|_{\mathrm{tr}} \leq t} \ \mathrm{trace}(\nabla f(X^{(k-1)})^\top S).$$

In class, it was claimed that $S^{(k-1)} = -t \cdot uv^\top$ where u, v are the leading left, right singular vectors of $\nabla f(X^{(k-1)})$. Prove this claim rigorously. *Note:* The trace of $A^\top B$ is the same as the inner product between vectorized matrices A and B and is equal to $\sum_{i,j} A_{i,j} B_{i,j}$.

5 Computational Problem (5 points)

We are getting close to the end of the semester and I would like to encourage you to work on your projects. So there is no computational problem on this homework. Instead, submit a 1-2 page description (typed up please, no handwritten scans allowed for this problem) of any preliminary computational results from your project. It is OK to report something you tried but did not work. It is also OK for project team members to submit the same answer to this problem (but not other problems!).