EE364b Spring 2023 Homework 7

Due Sunday 5/28 at 11:59pm via Gradescope

7.1 (8 points) As part of this homework, you will submit short reviews of the submitted project progress reports. On Canvas, we will assign two reports per student. You can find the projects assigned under the same assignment tab on Canvas. Point out the strengths and weaknesses of the proposed approaches. Please try to give constructive feedback.

- Please read the reports in-depth. Summarize the entire report in a short paragraph.
- Please try to give constructive feedback. Point out the strong and weak points. The following are possible points you can comment on: writing style, clarity, technical soundness and experimental evaluation (when applicable).
- You can suggest alternative approaches, e.g., use of different optimization methods, ideas or techniques to try out in that particular problem, any references you know that will be helpful, or any other application areas you find relevant.
- You can ask questions about anything that is missing or not clear in the report and provide comments regarding how to fix the issues. These questions and comments will be addressed in the final report.
- Finally, you will assign a tentative score to each midterm report based on two criteria: clarity/organization and technical content. Please see the grading rubric in the Canvas folder '`Files/Project/Grading`', and assign a score from 1 to 5 for each criteria (i.e., two scores per report) with 1 being the lowest and 5 the highest. Provide the justification in your review.

The poster presentations will be on June 7 from 1:30pm to 3:00pm during class hours.

Please review your assignments as soon as possible and email Aaron Mishkin immediately if you were assigned the same report twice or if you were assigned your own report. This is possible because some groups submitted multiple reports, one for each member.

After completing the reviews, you can submit them through the assignment "Midterm Progress Report" in the assignment section on Canvas.

Solution: This task is graded according to a binary scale (0:Incomplete, 1:Complete).

7.2 *(6 points) The Group Lasso and Interior Point Methods*
Consider again the group Lasso problem from Problem Set 5. Let $X \in \mathbb{R}^{n\times(p+1)}$ be a

data matrix where each column represents a predictor. Suppose that the matrix X is split into J groups over its columns:

$$X = [\vec{1}\ X_{(1)}\ X_{(2)} \dots\ X_{(J)}]$$

where $\vec{1} = [1\ 1 \dots 1] \in \mathbb{R}^n$ is a vector of all ones. The groups are typically determined by the types of predictors. To achieve sparsity over the groups rather than individual predictors, we may write $\beta = (\beta_0, \beta_{(1)}, \dots, \beta_{(J)})$, where β_0 is an intercept term and each $\beta_{(j)}$ is an appropriate coefficient block of β corresponding to $X_{(j)}$, and solve the regularized optimization problem:

$$\min_{\beta \in \mathbb{R}^{p+1}} f(\beta) + h(\beta).$$

Here $h(\beta)$ is a convex regularization term to promote the sparsity over groups. In this problem, we will use group Lasso to predict the Parkinson's disease (PD) symptom score on the Parkinsons dataset[1]. The PD symptom score is measured on the unified Parkinson's disease rating scale (UPDRS). This data contains 5, 785 observations, 18 predictors (provided in **X_train.csv**), and an outcome - the total UPDRS (provided in **y_train.csv**)

The 18 columns in the predictor matrix have the following groupings (in column ordering):

- age: Subject age in years
- sex: Subject gender, 0–male, 1–female
- Jitter(%), Jitter(Abs), Jitter:RAP, Jitter:PPQ5, Jitter:DDP: Several measures of variation in fundamental frequency of voice
- Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:APQ11, Shimmer:DDA: Several measures of variation in amplitude of voice
- NHR, HNR: Two measures of ratio of noise to tonal components in the voice
- RPDE: A nonlinear dynamical complexity measure
- DFA: Signal fractal scaling exponent
- PPE: A nonlinear measure of fundamental frequency variation

We consider the group Lasso problem, where $h(\beta) = \lambda \sum_j w_j \|\beta_{(j)}\|_2$:

$$\min_{\beta \in \mathbb{R}^{p+1}} \frac{1}{2n} \|X\beta - y\|_2^2 + \lambda \sum_j w_j \|\beta_{(j)}\|_2$$

A typical choice for weights on groups w_j is $\sqrt{p_j}$, where p_j is number of predictors that belong to the jth group, to account for the group sizes. We will solve the problem using an interior point method and the R-FISTA method from Problem Set 5.

[1] 'Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection', Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. BioMedical Engineering Online 2007, 6:23 (26 June 2007)

(a) (1 point) There are multiple ways to solve the group Lasso using interior point methods. First we consider reformulating the objective using second-order cone constraints. In particular, we introduce new variables $\alpha_{(i)}$ and consider solving the equivalent constrained problem

$$\min_{\beta,\alpha} \frac{1}{2n}\|X\beta - y\|_2^2 + \lambda \sum_j w_j \alpha_{(j)} \quad \text{s.t.} \quad \left\|\beta_{(j)}\right\|_2 \leq \alpha_{(j)},\ \forall j \tag{1}$$

Prove that the log-barrier function for the second order cone constraint $\left\|\beta_{(j)}\right\|_2 \leq \alpha_{(j)}$ given by

$$h(\beta_{(i)}, \alpha_{(i)}) = -\log(\alpha_{(i)}^2 - \left\|\beta_{(i)}\right\|_2^2),$$

is convex.

(b) (2 bonus points) Let $g(t) = h(\beta_{(i)} + tv, \alpha_{(i)} + tv)$ for some (u, v). Verify that the log-barrier h is self-concordant by proving that it satisfies satisfies the derivative inequality

$$|g'''(t)| \leq 2g''(t)^{3/2},$$

for any t and (u, v) such that $\left\|\beta_{(i)} + tu\right\|_2 \leq \alpha + tv$.

(c) (3 points) Implement the truncated Newton interior point method and use it to solve Eq. (1). You may use `scipy.sparse.linalg.cg` (or equivalent in another language) with the parameter `maxiter` to compute the truncated Newton direction. Set $\lambda = 0.02$ and use $\mu = 1.5$ to update the relaxation parameter t used by the interior point method. Initialize $t = 1$.

Hint: Start by using CVXPY to solve the centering problems and switch to truncated Newton method when you have verified the centering iterations are implemented properly.

(d) (1 point) Plot $f^k - f^*$ versus k for 50 iterations on a semi-log scale for the method in part (b) for the training data, where f^k denotes the objective value at step k, and the optimal objective value is $f^* = 49.9649$. Print the components of the solutions numerically. Which groups are selected, i.e., non-zero at the solution?

(e) (1 point) Compare the performance of this interior point method with R-FISTA from Problem Set 5. Which method would you prefer to use when solving this problem and why?