Assignment 1

CS6230: Optimization Methods in Machine Learning

Due: 26 Aug 2016, 13:00 hours
Max Marks: 100

INSTRUCTIONS

- Please submit your solutions to theory questions on an A4 sheet and hand it over to the TA (Adepu Ravi Sankar) in Room No 611, 6^{th} floor, Academic Block A by the due date (and time). Solutions to programming questions should be uploaded as a single ZIP file named '⟨YourRollNo⟩_assign1.zip' through the Google Classroom submission link.
- Your solution to coding problems should include plots and whatever explanation necessary to answer the questions asked.
- For late submissions, 10% is deducted for each day (including weekend) late after an assignment is due. Note that each student begins the course with 5 grace days for late submission of assignments. Late submissions will automatically use your grace days balance, if you have any left. You can see your balance on the CS6230 Marks and Grace Days document under the course Google drive.
- Please read the department plagiarism policy. Do not engage in any form of cheating - strict penalties will be imposed for both givers and takers. Please talk to instructor or TA if you have concerns.

1 THEORY (70 POINTS)

1. (1 point) What is the best optimization algorithm you have come across in your life and why?
2. (22 points) Prove or disprove the convexity of the following sets/functions:
   a. A *slab*, i.e a set of the form $\{x \in \mathbb{R}^n | \alpha \leq a^T x \leq \beta\}$ where $a \in \mathbb{R}^n, \alpha, \beta \in \mathbb{R}$
   b. A *wedge*, i.e $\{x \in \mathbb{R}^n | a_1^T x \leq \beta_1, a_2^T x \leq \beta_2\}$, where $a_1, a_2 \in \mathbb{R}^n, \beta_1, \beta_2 \in \mathbb{R}$
   c. The image of a convex set $S \subseteq \mathbb{R}^n$ under an affine transformation, i.e. the set $f(S)$ for $f(x) = Ax + b$, where $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$
   d. The set of points closer to a given point $x_0 \in \mathbb{R}^n$ than a given set $S \subseteq \mathbb{R}^n$, i.e. $\{x \in \mathbb{R}^n | ||x - x_0||_2 \leq ||x - y||_2, y \in S\}$
   e. The set of points closer to one set S than another set T, i.e $\{x \in \mathbb{R}^n | dist(x, S) \leq dist(x, T)\}$, where $S, T \subseteq \mathbb{R}^n$ and the distance from a point to a set is defined as $dist(x, S) := inf_{z\in S} ||x - z||_2\}$
   f. $f(x) = ||x||_p, x \in \mathbb{R}^d$ (Consider all possibilities of $p \in \mathbb{R}$)
   g. $g(x) = \sup_{y\in A} f(x, y)$ where $f(x, y)$ is convex in x for each $y \in A$

h. The sets $\{p(x)|p(x)$ is a probability distribution function$\}$ and $\{F(x)|F(x)$ is a cumulative distributive function$\}$.

i. Set of all PSD matrices

j. Trace(XX^T)

k. $f(w) = \sum_{i=1}^{n} \log p(y_i|w, x_i) + \frac{\lambda}{2}||w||^2, w \in \mathbb{R}^n, \lambda \geq 0$

3. (3 points) Draw the Venn Diagram for the sets Affine, Convex, Concave, Quasi-Convex, Strongly Convex, Strictly Convex

4. (5 points) List any five ways/approaches of proving a function is convex (A one-line explanation is sufficient for each method).

5. (5 points) Let $C \subseteq \mathbf{R}^n$ be the solution set of a quadratic inequality,

$$C = \{x \in \mathbf{R}^n | x^T Ax + b^T x + c \leq 0\},$$

with $A \in \mathbf{S}^n, b \in \mathbf{R}^n, c \in \mathbf{R}$,

a. Show that C is convex if $A \succeq 0$.

b. Is the converse of above statement true? Justify.

6. (8 points) Determine whether the following functions are *Lipschitz continuous* in the specified domain. Compute the *Lipschitz* constant if the function is *Lipschitz continuous* , otherwise explain why it is not *Lipschitz continuous.*

a. $f(x) = x, x \in [-1, 1]$

b. $f(x) = |x|, x \in \mathbb{R}$

c. $\sin x$

d. $f(x) = x^2 + |x|, x \in \mathbb{R}$

7. (5 points) Prove that if a function $f(x)$ is differentiable at $x^* \in (a, b)$ and has continuous derivative in $[a, b]$, then $f(x)$ is Lipschitz in $(x^* - \delta, x^* + \delta)$ for some $\delta > 0$.

8. (5 points) Show that every Lipschitz function is continuous.

9. (9 points) Let f be convex and twice differentiable, show that the following statements are equivalent:

- ∇f is Lipschitz with constant L
- $\nabla^2 f(x) \preceq LI$ for all x
- $f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}||y - x||_2^2$ for all x, y.

10. (7 points) Prove that *gradient descent* with fixed step size $t \leq 1/L$ satisfies

$$f(x^{(k)}) - f(x^*) \leq \frac{||x^{(0)} - x^*||^2}{2tk}$$

where k is the current iteration, $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable and additionally ∇f is Lipschitz continuous with constant $L > 0$.
Bonus (5 points): Comment about how the convergence criteria changes if the step size is not fixed.

2 PROGRAMMING (30 POINTS)

1. (12 points) CVX is a fantastic framework for disciplined convex programming-it's rarely the fastest tool for the job, but it's widely used, and so it's a great tool to be comfortable with. In this exercise, we will set up the CVX environment and solve a convex optimization problem. CVX variants are available for each of the major numerical programming languages. There are some minor syntactic and functional differences between the variants but all provide essentially the same functionality. The Matlab version (and by extension, the R version which calls Matlab under the covers) is the most mature but all should be sufficient for the purposes of this class. Download the CVX variant of your choosing:

   - Matlab - http://cvxr.com/cvx/
   - Python - http://www.cvxpy.org/en/latest/
   - Julia - https://github.com/JuliaOpt/Convex.jl
   - R - http://faculty.bscb.cornell.edu/~bien/cvxfromr.html

   and consult the documentation to understand the basic functionality.

   a. Using CVX, solve the 1d fused lasso problem

   $$\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^{n} (y_i - \beta_i)^2 + \lambda \sum_{i=1}^{n-1} |\beta_i - \beta_{i+1}|$$

   The data for this prob `y.txt` and `beta0.txt` are provided with this assignment.

   b. Load the data `y.txt` and solve the 1d fused lasso problem with $\lambda = 1$. Report the objective value obtained at the solution.

   c. Next, we consider how the solution changes as we vary λ. Solve the optimization problem for 100 logarithmically spaced values from 10^1 to 10^{-2} (in Matlab, `logspace(1, -2, 100)`). For each λ, compute the mean squared error (MSE) of the solution $\hat{\beta}$ and the true β_0 (from `beta0.txt`) as well as the number of change points in $\hat{\beta}$. For numerical purposes, we define a *changepoint* to be absolute difference greater than 10^{-8}. Plot MSE and number of changepoints as a function of λ.

2. (18 points) Disciplined Convex Programming (DCP) is a system for composing functions while ensuring their convexity. It is the language that underlies CVX. Essentially, each node in the parse tree for a convex expression is tagged with attributes for curvature (convex, concave, affine, constant) and sign (positive, negative) allowing for reasoning about the convexity of entire expressions. The website `http://dcp.stanford.edu/` provides visualization and analysis of simple expressions.

   Typically, writing problems in the DCP form is natural, but in some cases manipulation is required to construct expressions that satisfy the rules. For each set of mathematical expressions below (all define a convex set), give equivalent DCP expressions along with a brief explanation of why the DCP expressions are equivalent to the original. DCP expressions should be given in a form that passes analysis at `http://dcp.stanford.edu/analyzer`.

   a. $||(x, y, z||_2^2 \leq 1$
   b. $\sqrt{x^2 + 1} \leq 3x + y$
   c. $\frac{1}{x} + \frac{2}{y} \leq 5; x, y > 0$
   d. $(x + y)^2 / \sqrt{y} \leq x - y + 5, y > 0$
   e. $||(x + 2y, x - y)||_2 = 0$
   f. $log(e^{y-1} + e^{x/2} \leq -e^x$