convex optimization: formulation

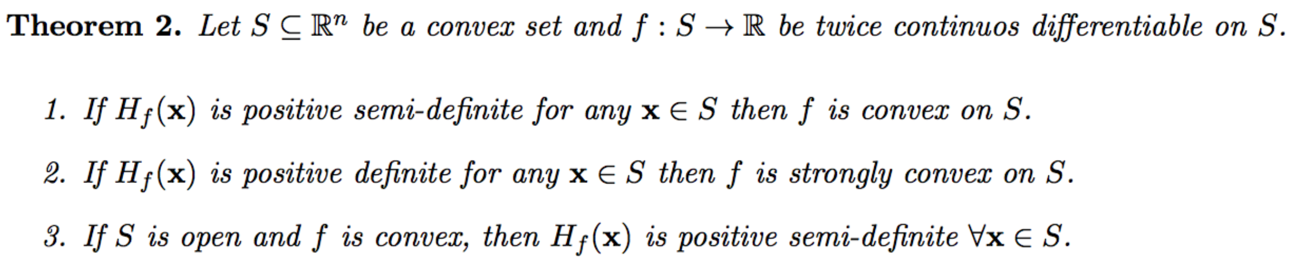
- optimization problems are everywhere in statistics and machine learning. we frame things in ml as an optimization problem P and we want to solve P. this class is about how to solve P. eg. in least squares regression, maximum likelihood estimation, traveling salesman, forms of motion planning, neural network training (some parts of it) and the list goes on..
- optimization problems are such we can study properties of the solutions to these problems without knowing exactly how we got there or without knowing what algorithm to use to get there. this is different from procedural things in stats/ml like hypothesis testing, boosting etc where the algorithm determines what we get to and we need to understand the algorithm to understand properties of the solutions.
- an important reason to study optimization algorithms is because different optimization algorithms work well to different degrees on different problems and on different kinds of data. understanding common algorithms, their properties and when they work well will help us choose the right algorithm/come up with new algorithms/modifications for our use case.
- another reason is because optimal solutions to problems might give us certain insights into the properties of the problem/properties of an optimal solution that is useful in itself even if we don't care about how we got to that solution.
- the right distinction between optimization problems used to be linear programming and non-linear programming. but it is now widely accepted the better way to slice them is convex problems vs non-convex problems. ryan doesn't give any reason why this is the case.
- "convex problems have provable solutions. all algorithms will work eventually if you run them long enough". non-convex problems are, in general, harder. the ways to solve them are not immediately obvious.
- convex sets are such that given any two x, y if we take $tx + (1-t)y$, that point should lie in the set for all $0 \leq t \leq 1$. non-convex sets are sets for which this is not true.
- domain of a function is the set of all values for which the function is defined and the value is finite.
- a function R^n → R is convex if:
    - its domain is a convex set
    - AND
    - the value of the function evaluated at a point between any two points is always less than the value of the function at each of the points.
      $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$, for all $0 \leq t \leq 1$ for all x, y in domain of f

  The condition on the domain being a convex set is to ensure that $f(tx + (1-t)y)$ is defined.
- a function f is concave if $-f$ is convex.
- a general optimization problem is one that involves finding $min f(x), x \epsilon D$ subject to $g_i(x) < 0$ and $h_j(x) = 0$ (basically some inequality and equality constraints) and D is the intersection of the domains of f, g, h.
- A convex optimization problem is an optimization problem where f, g_i are all convex and h_j is affine (in this class, affine basically just means linear). so $h_j(x) = a_j^T x + b_j$
- For convex optimization problems, local minima are global minima. local minima means it is feasible there (satisfies all the constraints) and it is lesser than all points in its neighborhood that are feasible. proof:

let x be a local minimum. lets say we know a z such that $f(z) < f(x)$ and $||z - x||^2 > p \rightarrow z$ is not in the neighborhood in which x is a local minimum. lets say $y = tx + (1 - t)z$. we can show that y is feasible by applying the definition of convexity and noticing that h is linear. so y is feasible and in the domain. we will choose t so that y lies in the neighborhood of x in which it is minimum. now $f(y) = f(tx + (1-t)z) \leq tf(x) + (1-t)f(z) < f(x)$, which is a contradiction. QED.

- earlier we defined convex sets in terms of two points. we are going to extend that definition to require all linear combinations with weights that add up to 1. the convex hull of a set C is $conv(C)$ is all linear combinations of elements with the above weight criterion.
- examples of convex sets: norm ball $x : ||x|| < r$, hyper plance $x : a^T x + b = 0$ for given a, b, half space $a^T x \leq b$, affine space $Ax = b$ (an affine space is a set of all solutions to a linear system of equations)
- Given a convex function $f : R^n \rightarrow R$ and a real number $\alpha \in R$, the α-sublevel set is defined as $x \epsilon D(f) : f(x) \leq \alpha$
- special cases:
  - We say that a convex optimization problem is a linear program if both the objective function and the constraints are linear
  - We say that a convex optimization problem is a quadratic program if both the objective function is quadratic but the the constraints are linear
  - We say that a convex optimization problem is a quadratically constrained quadratic program if both the objective function and the constraints are quadratic.

  In all the above cases, of course the objective function f is assumed to be convex.
- Note that we touched upon a proof idea for taylor expansion for multivariable functions in the multivariable calculus course. If we define the Hessian matrix, and take that proof a bit further it can be, with more manipulation, shown that:



Theorem 2. *Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \rightarrow \mathbb{R}$ be twice continuos differentiable on S.*

1. *If $H_f(\mathbf{x})$ is positive semi-definite for any $\mathbf{x} \in S$ then f is convex on S.*
2. *If $H_f(\mathbf{x})$ is positive definite for any $\mathbf{x} \in S$ then f is strongly convex on S.*
3. *If S is open and f is convex, then $H_f(\mathbf{x})$ is positive semi-definite $\forall \mathbf{x} \in S$.*

- The proof of the above is here. This taylor expansion, hessian and convexity condition will be useful all over the place.
- There is more notes on this about Lagrange multipliers, KKT optimality conditions and the use of this theory to derive a SVM algorithm in both the CS 229 Machine Learning notes and the notes for the Deep Learning book.