Topics in Machine Learning Midterm 1

2017-02-21

- A **mathematical optimization problem** is defined as: minimize $f_0(x)$ but subject to $f_i(x) \leq b_i, i = 1, \ldots, m$, where x is the optimization variable, f_0 is the **objective function**, f_i are the **constraint functions**.
- A **solution method** for a class of optimization problems is an algorithm that computes a solution of the problem (to some given accuracy), given a particular problem from the class, i.e., an **instance** of the problem.
- **Regularization** comes up in statistical estimation when the vector x to be estimated is given a prior distribution.
- Some optimization problems are considered to be easy ones because they have either a direct **analytical solution** or an efficient algorithm to solve them. These problems include least-squares, linear programming, and **convex optimization**. A **global** optimal is achievable.
- **Nonconvex optimization** sometimes has an analytical solution as well, e.g. in certain constraints. A global optimal is not guaranteed.
- The least-squares problem and linear programming problem are both special cases of the general convex optimization problem.
- Linear combination:
    - **Affine combination**: $\{\theta_1 x_1 + \ldots + \theta_k x_k | x \in C, \theta_1 + \ldots + \theta_k = 1\}$
    - **Conic combination**: $\{\theta_1 x_1 + \ldots + \theta_k x_k | x \in C, \theta \geq 0\}$
    - **Convex combination**: $\{\theta_1 x_1 + \ldots + \theta_k x_k | x \in C, \theta \geq 0, \theta_1 + \ldots + \theta_k = 1\}$
- **Affine**:
    - **Affine set**: the set C where the line through any two distinct points in C lies in C.
    - **Affine hull**: the set of all affine combinations of points in some set C. It is the smallest affine set that contains C.
- **Conic**:
    - **Conic hull**: the set of all conic combinations of points in C.
- **Convex**:
    - **Convex set**: the set C where the line segment between any two points in C lies in C.
    - **Convex hull**: the set of all convex combinations of points in C. It is the smallest convex set that contains C.

- Every affine set and cone is also convex. A cone is often called a **convex cone**.
- **Interior** v.s. **relative interior**
  - x is an **interior** point of C if there exists an open ball with the same dimension as C centered at x which is completely contained in C.
  - x is an **relative interior** point of C if there exists an open ball with the same dimension as the affine set of x centered at x which is completely contained in C.
- **Hyperplanes**: $\{x|a^{\mathrm{T}}x = b\}$ or $\{x|a^{\mathrm{T}}(x - x_0) = 0\}$
  - The solution set of a nontrivial linear equation among the components of x (and hence an affine set).
  - The set of points with a constant inner product to a given vector a.
  - A hyperplane with normal vector a and the offset b of the hyperplane from the origin.
  - An offset x_0, plus all vectors orthogonal to the (normal) vector a.
- **Polyhedra**: $\{x|Ax \preceq b, Cx = d\}$
  - The solution set of a finite number of linear equalities and inequalities.
  - The intersection of finite number of **hyperplanes** and **halfspaces**.
- **Ellipsoid**:
  - $\{x|(x - x_c)^{\mathrm{T}}P^{-1}(x - x_c) \leq 1\}$, where P is symmetric and positive definite.
  - $\{x_c + Au| \parallel u \parallel_2 \leq 1\}$, where a is square and nonsingular.
- **Norm ball**:
  - $\{x| \parallel x - x_c \parallel \leq r\}$
  - $\{x_c + ru| \parallel u \parallel \leq 1\}$
- **Norm cone**: $\{(x, t)| \parallel x \parallel \leq t\}$
- **Halfspaces**: $\{x|a^{\mathrm{T}}x \leq b\}$ or $\{x|a^{\mathrm{T}}(x - x0) \leq 0\}$
- **Simplexes**: the set of convex combinations of $k + 1$ affinely independent points in K dimensional space.
- **Positive semidefinite cone**: $S_+^n = \{X \in S^n|X \succeq 0\}$
  - S^n is a vector space with dimension $n(n + 1)/2$.

2017-03-07

- Operations that preserve convexity includes **intersection**, **affine functions**, **perspective functions**, and **linear-fractional functions**.
- **Intersection**:
  - The intersection of an infinite number of halfspaces.
- **Affine functions**: $f(x) = Ax + b$

    - A sum of a linear function and a constant.
    - The *image* and the *inverse image* of S under f.
    - Scaling, translation, or projection of a convex.
    - The sum, partial sum, direct or Cartesian product of convexes.
    - Polyhedron under affine function $f(x) = (b - Ax, d - Cx)$ is the Cartesian product of the nonnegative orthant and the origin $\{x|f(x) \in \mathbb{R}_+^n \times \{0\}\}$.
    - Solution set of linear matrix inequality under affine function $f(x) = B - A(x)$ is a positive semidefinite cone S_+^n.
    - Hyperbolic cone $\{x|x^{\mathrm{T}}Px \leq (c^{\mathrm{T}}x)^2, c^{\mathrm{T}}x \geq 0\}$ under affine function $f(x) = (P^{1/2}x, c^{\mathrm{T}}x)$ is a second-order cone $\{(z,t)|z^{\mathrm{T}}z \leq t^2, t \geq 0\}$.
    - Ellipsoid $\{x|(x - x_c)^{\mathrm{T}}P^{-1}(x - x_c) \leq 1\}$ under affine function $g(x) = P^{-1/2}(x - x_c)$ is a unit Euclidean ball $\{u| \parallel u \parallel \leq 1\}$.
- **Proper cone**: K
    - K is *convex*.
    - K is *closed*.
    - K is *solid*, which means it has nonempty interior.
    - K is *pointed*, which means that it contains no line.
- Examples of proper cones:
    - Nonnegative orthant and componentwise inequality.
    - Positive semidefinite cone and matrix inequality.
    - Cone of polynomials nonnegative on [0,1].
- A proper cone K can be used to define a *generalized inequality*, which is a partial ordering on $\mathbb{R}^n$.
- **Generalized inequality**: $x \preceq_K y \leftrightarrow y - x \in K$.
- $x \in S$ is the **minimum** element if and only if $S \subseteq x + K$, which means *all other points of S lie above and to the right*.
- $x \in S$ is the **minimal** element if and only if $(x - K) \cap S = x$, which means *no other point of S lies to the left and below x*.
- **Separating hyperplane theorem**: Suppose C and D are nonempty disjoint convex sets, i.e., $C \cap D = \emptyset$. Then there exist $a \neq 0$ and b such that $a^{\mathrm{T}}x \leq b\, \forall x \in C$ and $a^{\mathrm{T}}x \geq b\, \forall x \in D$. The hyperplane $\{x|a^{\mathrm{T}}x = b\}$ is called a **separating hyperplane** for the sets C and D.

2017-03-14

- **Supporting hyperplane theorem**: if C is convex, then there exists a supporting hyperplane at every boundary point of C.

- **Convex functions**: f is convex if $\mathbf{dom} f$ is a convex set and $f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$.
- **Concave functions**: f is concave if $-f$ is convex.
- **Affine functions**:
    - All affine (and therefore also linear) functions are both convex and concave.
    - Any function that is convex and concave is affine.
- Practical methods for establishing convexity of a function
    - Verify definition.
    - For first-order conditions, show $f(y) \ge f(x) + \nabla f(x)^{\mathrm{T}}(y-x)$
    - For twice differentiable functions, show its second derivative ∇^2 is greater than 0.
- First-order conditions:
    - For a convex function, the first-order Taylor approximation is in fact a global underestimator of the function.
    - If the first-order Taylor approximation of a function is always a global underestimator of the function, then the function is convex.
- Second-order conditions: **Hessian** (second derivative) test.
- Examples of convex functions: *affine*, *exponential*, *powers* (power ≥ 1, or power ≤ 0), *powers of absolute value*, *negative entropy*, *norms*, *spectral (maximum singular value) norm*, *max*, *quadratic*, *least-squares*, *quadratic-over-linear*, *log-sum-exp*, *geometric mean*, *log-determinant*, *Jensen's inequality*, etc.
- Examples of concave functions: *affine*, *powers* (0 ≤ power ≤ 1), *logarithm*, etc.

2017-03-21

- Operations that preserve convexity: *nonnegative weighted sum*, *composition with affine function*, *pointwise maximum*, *pointwise supremum*, *composition with scalar functions*, *vector composition*, *minimization*, *perspective function*, *conjugate function*, etc.
- **Nonnegative weighted sum** & **composition with affine function**:
    - Log barrier for linear inequalities
    - Any norm of affine function
- **Pointwise maximum**:
    - Piecewise-linear function
    - Sum of r largest components of $x \in \mathbb{R}^n$
- **Pointwise supremum**: $g(x) = \sup_{y \in C} f(x, y)$ is convex if $f(x, y)$ is convex in x for each $y \in C$
    - **Support function** of a set C: $S_C(x) = \sup_{y \in C} y^{\mathrm{T}} x$
    - Distance to farthest point in a set C: $f(x) = \sup_{y \in C} \| x - y \|$

- Maximum eigenvalue of symmetric matrix S: $\lambda_{\max}(S) = \sup_{\|y\|_2=1} y^{\mathrm{T}} S y$

- **Minimization**: $g(x) = \inf_{y \in C} f(x, y)$ is convex if $f(x, y)$ is convex in (x, y) and C is a convex set
  - $f(x, y) = x^{\mathrm{T}} A x + 2x^{\mathrm{T}} B y + y^{\mathrm{T}} C y$ with $[A, B; B^{\mathrm{T}}, C] \succeq 0, C \succeq 0$
  - Distance to a set C: $\mathbf{dist}(x, C) = \inf_{y \in C} \| x - y \|$ if C is convex

- **Conjugate functions**: $f^*(y) = \sup(y^{\mathrm{T}} x - f(x))$ must be convex.
  - Conjugate function of negative logarithm $f(x) = -\log x$
  - Conjugate function of strictly convex quadratic $f(x) = (1/2) x^{\mathrm{T}} Q x$

- **Quasiconvex functions**: f is quasiconvex if $\mathbf{dom} f$ is convex and the sublevel sets $S_\alpha = \{x \in \mathbf{dom} f | f(x) \leq \alpha\}$ are convex for all α
  - $\sqrt{x}$ is quasiconvex on $\mathbb{R}$
  - $\mathbf{ceil}(x) = \inf\{z \in Z | z \geq x\}$ is quasilinear
  - $\log x$ is quasilinear on $\mathbb{R}_{++}$
  - $f(x_1, x_2) = x_1 x_2$ is quasiconcave on $\mathbb{R}^2_{++}$
  - Linear-fractional function is quasilinear
  - Distance ratio is quasiconvex

- Properties of *quasiconvex functions*:
  - **Modified Jensen inequality**: $f(\theta x + (1-\theta) y) \leq \max\{f(x), f(y)\}$
  - *First-order condition*: $f(y) \leq f(x) \Rightarrow \nabla f(x)^{\mathrm{T}} (y - x) \leq 0$

- **Log-concave and log-convex functions**: $f(\theta x + (1-\theta) y) \geq f(x)^\theta f(y)^{1-\theta}$
  - Powers x^a on $\mathbb{R}_{++}$ is log-convex for $a \leq 0$, log-concave for $a \geq 0$
  - Many common probability densities are log-concave.
  - Cumulative Gaussian distribution function Φ is log-concave.

- Properties of *log-concave functions*:
  - Twice differentiable f with convex domain is log-concave if and only if $f(x) \nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^{\mathrm{T}}$
  - Product of log-concave functions is log-concave
  - Sum of log-concave functions is not always log-concave
  - Integration of log-concave functions is log-concave