Machine Learning & Optimization Lecture 8

- Optimization as minimization
- Convex Sets, Functions; Convex Optim Probs
- Jensen's Inequality
- First & Second Order Characterizations of Convexity
- Examples
- Operations that preserve convexity
- Examples of Cvx Optim Problems
Convex Optimization

- We’ve seen how to convert ML problems to optim prob using RERM and MLE frameworks.

- We also know we only need to ensure

\[ R_n(f_T) \leq R_n(f_{\hat{T}}) + O(\varepsilon) \]

- In general, optimization is NP-Hard, so we restrict ourselves to convex optimization problems (for now).

\[ x^* = \arg \min_{x \in C} f(x) \]

Convex function

Convex set

Non-convex

Local minimum

Global minimum = local minimum

\[ x^2 \]
Why restrict to cux optim probs

- Very expressive model class
- They are generally tractable
- Local minima are global minima

- Many ML problems are cux:
  - GLMs fit using MLE gives cux optim probs
  - Many efficient algo exist, tailored for different settings
Convexity

We say a set $C$ is convex if whenever $x, y \in C$, the line segment $[x, y] = \{ \alpha \in [0, 1] : \alpha x + (1-\alpha)y \}$ is contained in $C$.

$C$ - convex

We can show that $C$ is convex iff for all $x_1, \ldots, x_k \in C$ and any numbers $\Theta_1, \ldots, \Theta_k$ s.t. $\Theta_i > 0$ for all $i$ and $\sum_{i=1}^{k} \Theta_i = 1$, the convex combination

$$\Theta_1 x_1 + \cdots + \Theta_k x_k \in C$$

$C$ - nonconvex
Given \( x_1, \ldots, x_k \), the set of all convex combinations of these points is called their convex hull.

**Observe:**

If \( C \) is a convex set and \( X \) is a r.v. taking values in \( C \), then \( EX \in C \).

**Examples of convex sets**

- empty set
- \( \mathbb{R} \), \( \mathbb{R}_+ \), \( \mathbb{R}^d \), rays, lines, line segments
- hyperplanes: \( \{ x \mid a^T x = b \} = H \)
  - if \( x \in y \) in this hyperplane, then
  \[
  a^T [\alpha x + (1-\alpha) y] = \alpha a^T x + (1-\alpha) a^T y = \alpha b + (1-\alpha) b = b
  \]
  so \( [x, y] \subseteq H \)
- half-spaces are convex \[ \frac{1}{2} \mathbb{X} | a^T x \leq b \] \[ \mathbb{H} \leq \]

fix \( x, y \in \mathbb{H} \leq \) and \( \alpha \in [0, 1] \), then

\[
a^T \left[ \alpha x + (1-\alpha) y \right] \\
= \alpha a^T x + (1-\alpha) a^T y \\
\leq \alpha b + (1-\alpha) b = b
\]

so \([x, y] \in \mathbb{H} \leq \)

- convex hull of an arbitrary set \( S \)

\[
\text{conv}(S) = \sum_{i=1}^{n} \theta_i x_i + \cdots + \theta_k x_k : x_i \in S, \theta_i \geq 0 \\
\text{and } \sum_{i=1}^{k} \theta_i = 1 \]

is a convex set. This is the smallest convex set containing \( S \).
- norm balls: \[ B(x, r) = \{ y \mid \|x - y\| \leq r \} \]

\[ B_{\| \cdot \|_2} (0, 1) \]

\[ B_{\| \cdot \|_1} (0, 1) \]

- polyhedrons: set of solutions to a finite number of linear equalities and inequalities

\[ \frac{1}{2} x : a_1^T x \leq b_1 \]
\[ a_2^T x \leq b_2 \]
\[ a_3^T x = b_3 \]
\[ \vdots \]

\[ \Rightarrow \text{equiv to } a_3^T x \leq b_3 \]
\[ (-a_3)^T x \leq -b_3 \]

\[ = \frac{1}{2} x : A x \leq b \]
where \[ A = \begin{bmatrix} a_1^T \\ \vdots \\ a_N^T \end{bmatrix} \] and \[ b = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} \]
- Intersection of arbitrary # of convex sets is convex

\[ \bigcap_{i \in I} C_i \text{ is convex if } C_i \text{ is convex for all } i \in I \]
Convex function

A function $f$ defined on a convex set $C$ is called convex iff it satisfies

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$$

for all $\alpha \in [0,1]$ and $x, y \in C$

Ex: $\frac{1}{x}$ is convex for $x \in \mathbb{R}_{++}$

Graphically: a function is convex if it lies under its chords.
Consider epigraph of $f$

$$\text{epi}(f) = \{ (x, z) : f(x) \leq z \}$$

Fact: $f$ is convex iff $\text{epi}(f)$ is convex (on the set $\mathbb{C}$)

Proof $f \text{cux} \Rightarrow \text{epi}(f) \text{cux}$

Assume $f$ is convex, and $(x, z)$ and $(y, w) \in \text{epi}(f)$ on $\mathbb{C}$. We need to show that if $\alpha \in [0, 1]$ then $(\alpha x + (1-\alpha)y, \alpha z + (1-\alpha)w) \in \text{epi}(f)$.

To see this note that

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha z + (1-\alpha)w.$$ 

so $(\alpha x + (1-\alpha)y, \alpha z + (1-\alpha)w) \in \text{epi}(f)$ as claimed.
If a function $f$ is convex on a convex set $C$, then any local minimum in $C$ is a global minimum.

Most important consequence of convexity:

- Strict convexity
- Not strictly convex
- Not strictly convex
- Convex
- Strictly convex

$\text{P}(x) + (1-x)\text{P}(y) \leq x\text{P}(x) + (1-x)\text{P}(y)$
Proof (contradiction)

Assume that $x$ is a local minimizer (there is a neighborhood of $x$ s.t. $f(x) \leq f(y)$ for any $y$ in that neighborhood) but it is not a global minimizer. This means there exists a $z \in C$ such that $f(z) < f(x)$.

By convexity, for an $\epsilon \in (0, 1]$, $y = (1-\epsilon)x + \epsilon z\,$ then we have by convexity that $f(y) \leq (1-\epsilon)f(x) + \epsilon f(z) < f(x)$ because $f(z) < f(x)$.

We've shown that we can a point arbitrarily close to $x$ that satisfies $f(y) < f(x)$. This violates the assumption that $x$ is a local minimizer. We conclude that it's not possible to be a local minimizer without also being a global minimizer in $C$. 
Fact: if $f$ is strictly convex on $C$ then the minimizer of $f$ on $C$ (if any exists) is unique.
First and Second-Order Characterizations of Convexity

1) First-order: if $f$ is differentiable, then $f$ is convex iff \( \forall x, y \in C \)

\[
f(y) \geq f(x) + \nabla f(x)^T (y - x)
\]

That is, $f$ is convex iff it lies above all of its tangent lines.
Ex \( f(x) = x^2 \) is convex on \( \mathbb{R} \)

Note \( \frac{df}{dx} = 2x \)

Check \( f(y) \geq f(x) + \frac{df(x)}{dx} (y-x) \) for all \( x,y \in \mathbb{R} \):

\[
\begin{align*}
\Rightarrow & \quad y^2 \geq x^2 + 2x(y-x) \\
\Rightarrow & \quad y^2 \geq 2xy - x^2 \\
\Rightarrow & \quad x^2 + y^2 - 2xy \geq 0 \\
\Rightarrow & \quad (x-y)^2 \geq 0
\end{align*}
\]

Conclude that \( f(x) = x^2 \) is convex on \( \mathbb{R} \).
2) Second-order condition: if \( f \) is twice-differentiable then \( f \) is convex iff
\[
\nabla^2 f(x) \succeq 0 \quad \text{(Hessian is positive semidefinite (PSD))}
\]

Recall the Hessian of a function \( f : \mathbb{R}^d \to \mathbb{R} \) is the matrix of second derivatives:
\[
\begin{bmatrix}
\nabla^2 f(x) \\
\end{bmatrix}
\]

\[
_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j} (x) = \frac{\partial^2 f}{\partial x_j \partial x_i} (x)
\]

so \( \nabla^2 f(x) \) is always symmetric.

Characterizations of PSD matrices \( M \) (\( M \succeq 0 \)):

1) \( M \) is symmetric and all of its eigenvalues are nonnegative.

2) \( M = CC^T \) for some matrix \( C \).
If \( f: \mathbb{R} \rightarrow \mathbb{R} \) then \( \nabla^2 f = [\nabla^2 f]_{11} = \frac{d^2 f}{dx^2} \)

**Ex**

\( f(x) = -\log x \) is convex on \( \mathbb{R}_+^* \)

\[ \frac{df}{dx} = -\frac{1}{x} = -x^{-1} \]

\[ \frac{d^2f}{dx^2} = \frac{1}{x^2} \geq 0 \text{ for } x \in \mathbb{R}_+^* \]

so by the 2nd-order condition, \( -\log x \) is convex on \( \mathbb{R}_+^* \)
Ex (next time) logsumexp(x) is convex on $\mathbb{R}^d$

-see Jensen's ineq first

-implies multiclass logistic regression is a convex optim problem