CS6301: Optimization in Machine Learning

Lecture 3 & 4: Convexity and Convex Optimization

Rishabh Iyer

Department of Computer Science
University of Texas, Dallas
https://sites.google.com/view/cs-6301-optml/home

January 22, 2020

Outline

- Recap from Previous Lecture
- Basics of Convexity: Convex Sets and Convex Functions
- Properties and Examples of Convex functions
- Basic Subgradient Calculus: Subgradients for non-differentiable convex functions
- Understanding the Convexity of Machine Learning Loss Functions
- Convex Optimization Problems

Recap From Previous Lecture

- Review of Notation: Vectors and Matrices
- Derivatives, Partial Derivatives, Gradients and Hessians
- Implementing Loss Functions and Gradients in Python.
- Assignment 1 was posted last week. The due date for this assignment is January 31st. This can be a time consuming assignment so please start early.
- Feel free to ask me any questions after class or during my office hours.
- **Hopefully all of you have started the assignments and are halfway through!**

Recap: Logistic Regression Gradient

- Lets start with Regularized Logistic Regression. Assume the Labels $y_i \in \{-1, +1\}$.
- The objective of Reg Logistic Loss is:

$$L(w) = \lambda/2\|w\|^2 + \sum_{i=1}^{n} \log(1 + \exp(-y_i w^T x_i)) \tag{1}$$

- Compute the gradient of this Loss?
- Gradient:

$$\begin{aligned}\nabla L(w) &= \lambda w + \sum_{i=1}^{n} \frac{-y_i \exp(-y_i(w^T x_i))}{1 + \exp(-y_i w^T x_i)} x_i \\ &= \lambda w + \sum_{i=1}^{n} \frac{-y_i}{1 + \exp(y_i w^T x_i)} x_i\end{aligned}$$

Recap: Logistic Regression Hessian

- Lets next compute the Hessian.
- Recall the Gradient:

$$\nabla L(w) = \lambda w + \sum_{i=1}^{n} \frac{-y_i}{1 + \exp(y_i w^T x_i)} x_i$$

- You can derive the Hessian as:

$$\nabla^2 L(w) = \lambda I + \sum_{i=1}^{n} \frac{\exp(y_i w^T x_i)}{(1 + \exp(y_i w^T x_i))^2} x_i x_i^T$$

- Define $\sigma(z) = 1/(1 + \exp(-z))$. Then its easy to see that:

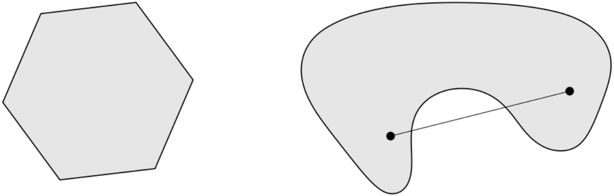
$$\nabla^2 L(w) = \sigma(y_i w^T x_i)(1 - \sigma(y_i w^T x_i)) x_i x_i^T + \lambda I$$

Numerical Issues and Implementations

- We studied numerical issues with $\log(1 + \exp(-x))$
- How do we fix it? See question in Assignment 1!
- Also numerical issues with $\log(\exp(x1) + \exp(x2))$ or $\exp(x1)/(\exp(x1) + \exp(x2))$
- We also covered how to implement the Logistic Loss Function.

Convex Sets

A set C is a **convex set** if the line segment between any two points of C lies in C, i.e. if for any $x, y \in C$ and for any $0 < \lambda < 1$, we have that $\lambda x + (1 - \lambda) y \in C$.



Source: Boyd's Textbook

Properties of Convex Sets

- Intersections of Convex Sets are Convex. Let $C_1, \cdots, C_k$ be convex sets, then $\cap_{i=1}^k C_i$ is convex.
- Is the union of convex sets convex?
- Projections onto convex sets are unique (and often efficient to compute).

$$P_C(x) = \text{argmin}_{y \in C} \|y - x\|$$

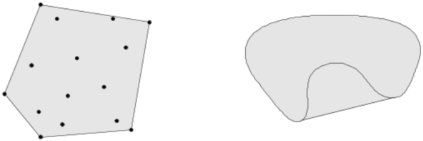
- Examples of Convex Sets:
  - $C = \{x \in \mathbb{R}^n : \|x\| \leq k\}$
  - $C = \{x \in \mathbb{R}^n : w^T x \leq k\}$
  - Given a convex function f, the associated set $C_f = \{x \in \mathbb{R}^n : f(x) \leq k\}$ is convex.

Convex combination and convex hull

- **Convex combination** of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\begin{aligned} \mathbf{x} = & \quad \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + ... + \theta_k \mathbf{x}_k = conv(\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}) \\ \text{with} & \quad \theta_1 + \theta_2 + ... + \theta_k = 1, \theta_i \geq 0. \end{aligned}$$

- **Convex hull or conv(S)** is the set of all convex combinations of point in the set S.



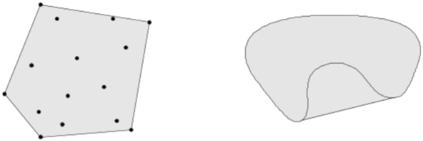
- Should S be always convex?
- What about the convexity of conv(S)?

Convex combination and convex hull

- **Convex combination** of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k$ is any point $\mathbf{x}$ of the form

$$\begin{aligned} \mathbf{x} = & \quad \theta_1 \mathbf{x}_1 + \theta_2 \mathbf{x}_2 + ... + \theta_k \mathbf{x}_k = conv(\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k\}) \\ \text{with} & \quad \theta_1 + \theta_2 + ... + \theta_k = 1, \theta_i \geq 0. \end{aligned}$$

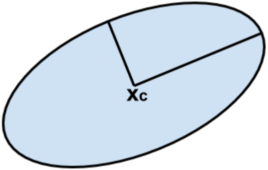
- **Convex hull or conv(S)** is the set of all convex combinations of point in the set S.



- Should S be always convex? No.
- What about the convexity of conv(S)? It's always convex.

Euclidean balls and ellipsoids

- **Euclidean ball** with **center** $\mathbf{x}_c$ and **radius** r is given by:
  $B(\mathbf{x}_c, r) = \{\mathbf{x} — \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\} = \{\mathbf{x}_c + ru — \|u\|_2 \leq 1\}$
- **Ellipsoid** is a **set** of form:
  $\{\mathbf{x} — (\mathbf{x} - \mathbf{x}_c)^T P^{-1}(\mathbf{x} - \mathbf{x}_c) \leq 1\}$, where P $\in S^n_{++}$ i.e. P is SPD matrix.
  - Other representation: $\{\mathbf{x}_c + \text{A u} — \|u\|_2 \leq 1\}$ with A square and non-singular(i.e. A^{-1} exists).



Norm balls

- **Recap Norm:** A function[1] $\|.\|$ that satisfies:
  1. $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$.
  2. $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$ for any scalar $\alpha \in \Re$.
  3. $\|\mathbf{x}_1 + \mathbf{x}_2\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$ for any vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.
- **Norm ball** with **center** $\mathbf{x}_c$ and **radius** r: $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_x\| \leq r\}$ is a convex set. Why?

[1] ($\|.\|$ is a general (unspecified) norm; $\|.\|_{symb}$ is particular norm.)

Norm balls

- **Recap Norm:** A function[1] $\|.\|$ that satisfies:
  1. $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$.
  2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for any scalar $\alpha \in \Re$.
  3. $\|\mathbf{x}_1 + \mathbf{x}_2\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$ for any vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.
- **Norm ball** with **center** $\mathbf{x}_c$ and **radius** r: $\{\mathbf{x} | \|\mathbf{x} - \mathbf{x}_x\| \leq r\}$ is a convex set. Why?
  - Eg 1: **Ellipsoid** is defined using $\|\mathbf{x}\|_P^2 = \mathbf{x}^T P \mathbf{x}$.
  - Eg 2: **Euclidean ball** is defined using $\|\mathbf{x}\|_2$.

[1] ($\|.\|$ is a general (unspecified) norm; $\|.\|_{symb}$ is particular norm.)

Convex Functions

- A Function $f : \mathbb{R}^d \to \mathbb{R}$ is convex if:
  - $dom(f)$ is a convex set
  - for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
- Geometrically, the line segment between $(x, f(x))$ and $(y, f(y))$ lies above the graph of f.

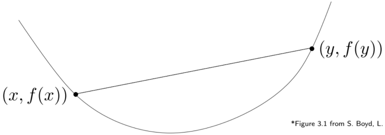
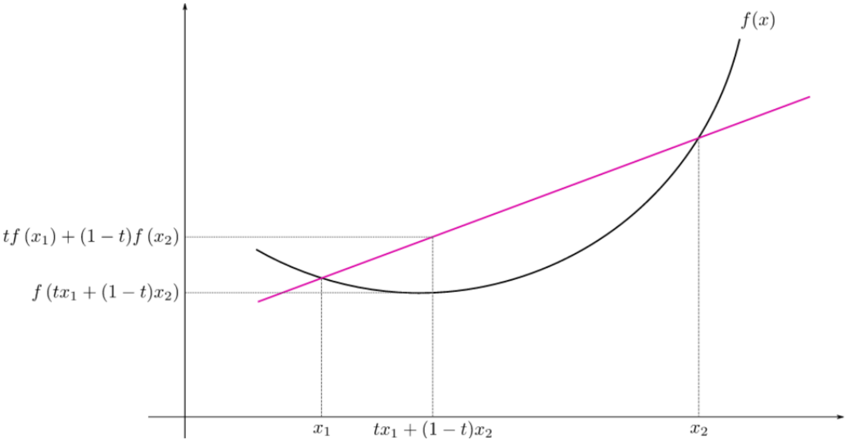


*Figure 3.1 from S. Boyd, L. Vandenberghe

- f is strictly convex if for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$

Intuition of Convexity



Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when *dom*(f) is a convex set:

1. f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$

Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when $dom(f)$ is a convex set:

1. f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
2. f is convex iff $\forall x_1, x_2, x_3$ such that $x_1 < x_2 < x_3$ it holds that $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$

Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when $dom(f)$ is a convex set:

1. f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$
2. f is convex iff $\forall x_1, x_2, x_3$ such that $x_1 < x_2 < x_3$ it holds that $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$
3. f is convex iff $f'(x)$ is a monotonic function of x. In other words, $f'(x_2) \geq f'(x_1)$ if $x_2 \geq x_1$.

Equivalent Definitions of Convex Functions

The following conditions are equivalent (in one dimension) when $dom(f)$ is a convex set:

1. f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
2. f is convex iff $\forall x_1, x_2, x_3$ such that $x_1 < x_2 < x_3$ it holds that $\frac{f(x_2)-f(x_1)}{x_2-x_1} \leq \frac{f(x_3)-f(x_2)}{x_3-x_2}$
3. f is convex iff $f'(x)$ is a monotonic function of x. In other words, $f'(x_2) \geq f'(x_1)$ if $x_2 \geq x_1$.
4. f is convex iff $f''(x) \geq 0$

Are the following functions convex?

- $f(x) = \exp(x)$
- $f(x) = \exp(-x)$
- $f(x) = \log x$
- $f(x) = \sin x$
- $f(x) = \log(1 + \exp(-x))$
- $f(x) = x^2$
- $f(x) = x^{2n}$ where n is an integer
- $f(x) = \max\{x, 0\}$
- $f(x) = \sqrt{x}$

From 1 dimensions to n dimensions

- Conditions for convexity in 1 dimensions is eas(ier)
- In the rest of this lecture, we shall understand how to extend this to n dimensions.
- Note that the basic definition of convexity still holds: f is convex iff for all $x, y \in dom(f)$ and $\lambda : 0 < \lambda < 1$, we have: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
- We shall look at some results which will help us prove some functions are convex!

Strongly Convex Functions

- A Function $f : \mathbb{R}^d \to \mathbb{R}$ is strongly convex if there exists a $\mu > 0$ such that the function $g(x) = f(x) - \mu/2\|x\|^2$ is convex
- The parameter μ is the strong convexity parameter
- Geometrically, strong convexity means that there exists a quadratic lower bound on the growth of the function.
- Its easy to see that Strong Convexity implies Strict Convexity!
- Strong Convexity Doesn't imply the function is differentiable!
- If a function f is strongly convex and g is convex (not necessarily strongly convex), $f + g$ is strongly convex.
- $\|x\|^2$ is strongly convex!
- Hence for any convex function f, the function $f(x) + \lambda/2\|x\|^2$ is strongly convex!
- To summarize: Strong Convexity $\Rightarrow$ Strict Convexity $\Rightarrow$ Convexity! (The converse does not hold)

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$
- Exponential: $f(x) = exp(\alpha x)$

Convex Functions

Examples of Convex Functions

- Linear Functions: $f(x) = a^T x$
- Affine Functions: $f(x) = a^T x + b$
- Exponential: $f(x) = exp(\alpha x)$
- Every Norm is Convex. **Why?**
  - By Triangle Inequality: $f(x+y) \leq f(x) + f(y)$, and homogeneity of norm: $f(\alpha x) = \alpha f(x)$ for a scalar α
  - It follows that

$$f(\lambda x + (1-\lambda)y) \leq f(\lambda x) + f((1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y)$$

Properties of Convex Functions

- **Non-negative weighted sum:** $f = \sum_{i=1}^{n} \alpha_i f_i$ is convex if each f_i for $1 \leq i \leq n$ is convex and $\alpha_i \geq 0, 1 \leq i \leq n$.

Properties of Convex Functions

- **Non-negative weighted sum:** $f = \sum_{i=1}^{n} \alpha_i f_i$ is convex if each f_i for $1 \leq i \leq n$ is convex and $\alpha_i \geq 0, 1 \leq i \leq n$.
- **Composition with Affine function:** $f(Ax + b)$ is convex if f is convex. For example:

Properties of Convex Functions

- **Non-negative weighted sum:** $f = \sum_{i=1}^{n} \alpha_i f_i$ is convex if each f_i for $1 \leq i \leq n$ is convex and $\alpha_i \geq 0, 1 \leq i \leq n$.
- **Composition with Affine function:** $f(Ax + b)$ is convex if f is convex. For example:
  - The log barrier for linear inequalities, $f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$, is convex since $-\log(x)$ is convex.

Properties of Convex Functions

- **Non-negative weighted sum:** $f = \sum_{i=1}^{n} \alpha_i f_i$ is convex if each f_i for $1 \leq i \leq n$ is convex and $\alpha_i \geq 0, 1 \leq i \leq n$.
- **Composition with Affine function:** $f(Ax + b)$ is convex if f is convex. For example:
  - The log barrier for linear inequalities, $f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$, is convex since $-\log(x)$ is convex.
  - Any norm of an affine function, $f(x) = ||Ax + b||$, is convex.

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing
- Proof idea: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing
- Proof idea: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).
- Examples:

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing
- Proof idea: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).
- Examples:
  - $f(x) = \exp(f(x))$ is convex if f is convex

Composition with Scalar Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.

$$f(x) = h(g(x))$$

- f is convex if a) g convex, h convex and non-decreasing or b) g concave, h convex and non-increasing
- Proof idea: Take double derivative and try to show that $\nabla^2 f \geq 0$ (easier to prove this for $m = 1$).
- Examples:
  - $f(x) = \exp(f(x))$ is convex if f is convex
  - $1/g(x)$ is convex if g is concave.

Composition with Vector Functions

- Composition of $g : \mathbb{R}^n \rightarrow \mathbb{R}^k$ and $h : \mathbb{R}^k \rightarrow \mathbb{R}$.

$$f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x))$$

Composition with Vector Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$.

$$f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x))$$

- f is convex if a) g_i's convex, h convex and non-decreasing in each argument or b) g_i concave, h convex and non-increasing in each argument

Composition with Vector Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$.

$$f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x))$$

- f is convex if a) g_i's convex, h convex and non-decreasing in each argument or b) g_i concave, h convex and non-increasing in each argument
- Examples:

Composition with Vector Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$.

$$f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x))$$

- f is convex if a) g_i's convex, h convex and non-decreasing in each argument or b) g_i concave, h convex and non-increasing in each argument
- Examples:
  - $f(x) = \sum_i \log(g(x))$ is concave if g is concave and positive

Composition with Vector Functions

- Composition of $g : \mathbb{R}^n \to \mathbb{R}^k$ and $h : \mathbb{R}^k \to \mathbb{R}$.

$$f(x) = h(g(x)) = h(g_1(x), \cdots, g_k(x))$$

- f is convex if a) g_i's convex, h convex and non-decreasing in each argument or b) g_i concave, h convex and non-increasing in each argument
- Examples:
  - $f(x) = \sum_i \log(g(x))$ is concave if g is concave and positive
  - $\log \sum_{i=1}^k \exp(g_i(x)$ is convex if g_i is convex.

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\right\}$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\right\}$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
    - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
    - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\left\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\right\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
  - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
    - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
    - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is

Pointwise Maximums and Supremums

Following functions are convex, but may not be differentiable everywhere.

- **Pointwise maximum:** If $f_1, f_2, \ldots, f_m$ are convex, then $f(\mathbf{x}) = max\{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})\}$ is also convex. For example:
  - Sum of r largest components of $\mathbf{x} \in \Re^n$ $f(\mathbf{x}) = x_{[1]} + x_{[2]} + \ldots + x_{[r]}$, where $x_{[1]}$ is the i^{th} largest component of $\mathbf{x}$, is a convex function.
- **Pointwise supremum:** If $f(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ for every $\mathbf{y} \in \mathcal{S}$, then $g(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{S}} f(\mathbf{x}, \mathbf{y})$ is convex. For example:
  - The function that returns the maximum eigenvalue of a symmetric matrix X, *viz.*, $\lambda_{max}(X) = \sup_{\mathbf{y} \in \mathcal{S}} \frac{\|X\mathbf{y}\|_2}{\|\mathbf{y}\|_2}$ is a convex function of the symmetrix matrix X.

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression:
  $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i)) + \lambda||\theta||$

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i)) + \lambda \|\theta\|$
- L1/L2 Reg SVMs: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\} + \lambda \|\theta\|$

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i)) + \lambda\|\theta\|$
- L1/L2 Reg SVMs: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i\theta^T x_i\} + \lambda\|\theta\|$
- L1/L2 Reg Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\} + \sum_{i=1}^{c} \lambda \sum_{j=1}^{m} \|\theta_j\|$

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i)) + \lambda \|\theta\|$
- L1/L2 Reg SVMs: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\} + \lambda \|\theta\|$
- L1/L2 Reg Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\} + \sum_{i=1}^{c} \lambda \sum_{j=1}^{m} \|\theta_j\|$
- L1/L2 Reg Least Squares (Lasso): $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2 + \lambda \|\theta\|$

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i\theta^T x_i)) + \lambda\|\theta\|$
- L1/L2 Reg SVMs: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i\theta^T x_i\} + \lambda\|\theta\|$
- L1/L2 Reg Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\} + \sum_{i=1}^{c} \lambda \sum_{j=1}^{m} \|\theta_j\|$
- L1/L2 Reg Least Squares (Lasso): $L(\theta) = \sum_{i=1}^{n}(\theta^T x_i - y_i)^2 + \lambda\|\theta\|$
- Matrix Completion: $L(X) = \sum_{i=1}^{n} ||y_i - A_i(X)||_2^2 + ||X||_*$

Which of the Following Loss Functions are Convex?

- L1/L2 Reg Logistic Regression: $L(\theta) = \sum_{i=1}^{n} \log(1 + \exp(-y_i \theta^T x_i)) + \lambda \|\theta\|$
- L1/L2 Reg SVMs: $L(\theta) = \sum_{i=1}^{n} \max\{0, 1 - y_i \theta^T x_i\} + \lambda \|\theta\|$
- L1/L2 Reg Multi-class Logistic Regression: $L(\theta_1, \cdots, \theta_k) = \sum_{i=1}^{n} -\theta_{y_i}^T x_i + \log(\sum_{c=1}^{k} \exp(\theta_c^T x_i))\} + \sum_{i=1}^{c} \lambda \sum_{j=1}^{m} \|\theta_j\|$
- L1/L2 Reg Least Squares (Lasso): $L(\theta) = \sum_{i=1}^{n} (\theta^T x_i - y_i)^2 + \lambda \|\theta\|$
- Matrix Completion: $L(X) = \sum_{i=1}^{n} ||y_i - A_i(X)||_2^2 + ||X||_*$
- Soft-Max Contextual Bandits: $L(\theta) = \sum_{i=1}^{n} \frac{r_i}{p_i} \frac{\exp(\theta^T x_i^{a_i})}{\sum_{j=1}^{k} \exp(\theta^T x_i^j)} + \lambda \|\theta\|$

The Direction Vector

- Consider a function $f(\mathbf{x})$, with $\mathbf{x} \in \Re^n$.
- We start with the concept of the direction at a point $\mathbf{x} \in \Re^n$.
- We will represent a vector by $\mathbf{x}$ and the k^{th} component of $\mathbf{x}$ by x_k.
- Let $\mathbf{u}^k$ be a unit vector pointing along the k^{th} coordinate axis in $\Re^n$;
- $u_k^k = 1$ and $u_j^k = 0, \ \forall j \neq k$
- An arbitrary direction vector $\mathbf{v}$ at $\mathbf{x}$ is a vector in $\Re^n$ with unit norm ($i.e.$, $||\mathbf{v}|| = 1$) and component v_k in the direction of $\mathbf{u}^k$.

Directional derivative and the gradient vector

Let $f : \mathcal{D} \rightarrow \Re$, $\mathcal{D} \subseteq \Re^n$ be a function.

Definition

[Directional derivative]: The *directional derivative* of $f(\mathbf{x})$ at $\mathbf{x}$ in the direction of the unit vector $\mathbf{v}$ is

Directional derivative and the gradient vector

Let $f : \mathcal{D} \to \Re$, $\mathcal{D} \subseteq \Re^n$ be a function.

Definition

[Directional derivative]: The *directional derivative* of $f(\mathbf{x})$ at $\mathbf{x}$ in the direction of the unit vector $\mathbf{v}$ is

$$D_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \tag{2}$$

provided the limit exists.

Directional Derivative

As a special case, when $\mathbf{v} = \mathbf{u}^k$ the directional derivative reduces to the partial derivative of f with respect to x_k.

$$D_{\mathbf{u}^k} f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_k}$$

If $f(\mathbf{x})$ is a differentiable function of $\mathbf{x} \in \Re^n$, then f has a directional derivative in the direction of any unit vector $\mathbf{v}$, and

$$D_{\mathbf{v}} f(\mathbf{x}) = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_k} v_k = \nabla f^T v \tag{3}$$

- Lets define *sub-level sets* of a convex function as follows:

Definition

[Sublevel Sets]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \to \Re$. The set

$$L_\alpha(f) = \{\mathbf{x} | \mathbf{x} \in \mathcal{D},\ f(\mathbf{x}) \leq \alpha\}$$

is called the $\alpha-$sub-level set of f.

Now if a function f is convex,

Sublevel Sets of Convex Functions

- Lets define *sub-level sets* of a convex function as follows:

Definition

[Sublevel Sets]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \to \Re$. The set

$$L_\alpha(f) = \{\mathbf{x} | \mathbf{x} \in \mathcal{D},\ f(\mathbf{x}) \leq \alpha\}$$

is called the $\alpha-$sub-level set of f.

Now if a function f is convex, its $\alpha-$sub-level set is a convex set.

Convex Function $\Rightarrow$ Convex Sub-level sets

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$ be a convex function. Then $L_\alpha(f)$ is a convex set for any $\alpha \in \Re$.

Proof: Consider $\mathbf{x}_1, \mathbf{x}_2 \in L_\alpha(f)$. Then by definition of the level set, $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, $f(\mathbf{x}_1) \leq \alpha$ and $f(\mathbf{x}_2) \leq \alpha$. From convexity of $\mathcal{D}$ it follows that for all $\theta \in (0,1)$, $\mathbf{x} = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{D}$. Moreover, since f is also convex,

$$f(\mathbf{x}) \leq \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \leq \theta\alpha + (1-\theta)\alpha = \alpha$$

which implies that $\mathbf{x} \in L_\alpha(f)$. Thus, $L_\alpha(f)$ is a convex set. $\square$

The converse of this theorem does not hold. To illustrate this, consider the function $f(\mathbf{x}) = \frac{x_2}{1+2x_1^2}$. The 0-sublevel set of this function is $\{(x_1, x_2) \mid x_2 \leq 0\}$, which is convex. However, the function $f(\mathbf{x})$ itself is not convex.

Convex Function $\Rightarrow$ Convex Sub-level sets

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$ be a convex function. Then $L_\alpha(f)$ is a convex set for any $\alpha \in \Re$.

Proof: Consider $\mathbf{x}_1, \mathbf{x}_2 \in L_\alpha(f)$. Then by definition of the level set, $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, $f(\mathbf{x}_1) \leq \alpha$ and $f(\mathbf{x}_2) \leq \alpha$. From convexity of $\mathcal{D}$ it follows that for all $\theta \in (0,1)$, $\mathbf{x} = \theta \mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{D}$. Moreover, since f is also convex,

$$f(\mathbf{x}) \leq \theta f(\mathbf{x}_1) + (1-\theta) f(\mathbf{x}_2) \leq \theta\alpha + (1-\theta)\alpha = \alpha$$

which implies that $\mathbf{x} \in L_\alpha(f)$. Thus, $L_\alpha(f)$ is a convex set. $\square$

The converse of this theorem does not hold. To illustrate this, consider the function $f(\mathbf{x}) = \frac{x_2}{1+2x_1^2}$. The 0-sublevel set of this function is $\{(x_1, x_2) \mid x_2 \leq 0\}$, which is convex. However, the function $f(\mathbf{x})$ itself is not convex.

A function is called quasi-convex if all its sub-level sets are convex

Convex Sub-level sets $\implies$ Convex Function

A function is called quasi-convex if all its sub-level sets are convex sets. Every quasi-convex function is not convex!

Consider the Negative of the normal distribution $-\frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

This function is quasi-convex but not convex.

Consider the simpler function $f(x) = -exp(-(x-\mu)^2)$.

- Then $f'(x) = 2(x-\mu)exp(-(x-\mu)^2)$
- And $f''(x) = 2exp(-(x-\mu)^2) - 4(x-\mu)^2 exp(-(x-\mu)^2) = (2-4(x-\mu)^2)exp(-(x-\mu)^2)$ which is < 0 if $(x-\mu)^2 > \frac{1}{2}$,
- Thus, the second derivative is negative if $x > \mu + \frac{1}{\sqrt{2}}$ or $x < -\mu - \frac{1}{\sqrt{2}}$.
- Recall from discussion of convexity of $f : \Re \to \Re$ if the derivative is not non-decreasing everywhere $\implies$ function is not convex everywhere.

To prove that this function is quasi-convex, we can ....

Proof that the function is Quasi-Convex

1. Inspect the $L_\alpha(f)$ sublevel sets of this function:
   $L_\alpha(f) = \{x | -exp(-(x-\mu)^2) \leq \alpha\} = \{x | exp(-(x-\mu)^2) \geq -\alpha\}$.
2. Since $exp(-(x-\mu)^2)$ is monotonically increasing for $x < \mu$ and monotonically decreasing for $x > \mu$, the set $\{x | exp(-(x-\mu)^2) \geq -\alpha\}$ will be a contiguous closed interval around μ and therefore a convex set.
3. Thus, $f(x) = -exp(-(x-\mu)^2)$ is quasi-convex (and so is its generalization - the negative of the normal density function).

- One can similarly prove that the negative of the multivariate normal density function is also quasi-convex, by inspecting its sub-level sets, which are nothing but **ellipsoids**.

Proof that the function is Quasi-Convex

1. Inspect the $L_\alpha(f)$ sublevel sets of this function: $L_\alpha(f) = \{x| - exp(-(x-\mu)^2) \leq \alpha\} = \{x|exp(-(x-\mu)^2) \geq -\alpha\}$.
2. Since $exp(-(x-\mu)^2)$ is monotonically increasing for $x < \mu$ and monotonically decreasing for $x > \mu$, the set $\{x|exp(-(x-\mu)^2) \geq -\alpha\}$ will be a contiguous closed interval around μ and therefore a convex set.
3. Thus, $f(x) = -exp(-(x-\mu)^2)$ is quasi-convex (and so is its generalization - the negative of the normal density function).

- One can similarly prove that the negative of the multivariate normal density function is also quasi-convex, by inspecting its sub-level sets, which are nothing but **ellipsoids**.

Convex Functions and Their Epigraphs

Let us further the connection between convex functions and sets by introducing the concept of the *epigraph* of a function.

Definition

[Epigraph]: Let $\mathcal{D} \subseteq \Re^n$ be a nonempty set and $f : \mathcal{D} \to \Re$. The set $\{(\mathbf{x}, f(\mathbf{x}) | \mathbf{x} \in \mathcal{D}\}$ is called graph of f and lies in $\Re^{n+1}$. The epigraph of f is a subset of $\Re^{n+1}$ and is defined as

$$epi(f) = \{(\mathbf{x}, \alpha) | f(\mathbf{x}) \leq \alpha, \ \mathbf{x} \in \mathcal{D}, \ \alpha \in \Re\} \tag{4}$$

In some sense, the epigraph is the set of points lying above the graph of f.

Eg: Recall affine functions of vectors: $\mathbf{a}^T\mathbf{x} + b$ where $\mathbf{a} \in \Re^n$. Its epigraph is $\{(\mathbf{x}, t) | \mathbf{a}^T\mathbf{x} + b \leq t\} \subseteq \Re^{n+1}$ which is a half-space (a convex set).

Convex Functions and Their Epigraphs (contd)

There is a one to one correspondence between the convexity of function f and that of the set $epi(f)$, as stated in the following result.

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \rightarrow \Re$. Then

Convex Functions and Their Epigraphs (contd)

There is a one to one correspondence between the convexity of function f and that of the set $epi(f)$, as stated in the following result.

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$. Then f is convex if and only if $epi(f)$ is a convex set.

Proof: f **convex function** $\implies$ $epi(f)$ **convex set**

Convex Functions and Their Epigraphs (contd)

There is a one to one correspondence between the convexity of function f and that of the set $epi(f)$, as stated in the following result.

Theorem

Let $\mathcal{D} \subseteq \Re^n$ be a nonempty convex set, and $f : \mathcal{D} \to \Re$. Then f is convex if and only if $epi(f)$ is a convex set.

Proof: f **convex function** $\implies$ $epi(f)$ **convex set**

Let f be convex. For any $(\mathbf{x}_1, \alpha_1) \in epi(f)$ and $(\mathbf{x}_2, \alpha_2) \in epi(f)$ and any $\theta \in (0, 1)$,

$$f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2)) \leq \theta\alpha_1 + (1-\theta)\alpha_2$$

Since $\mathcal{D}$ is convex, $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathcal{D}$. Therefore, $(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \theta\alpha_1 + (1-\theta)\alpha_2) \in epi(f)$. Thus, $epi(f)$ is convex if f is convex. This proves the necessity part.

Convex Functions and Their Epigraphs (contd)

$epi(f)$ convex set $\implies$ f convex function

To prove sufficiency, assume that $epi(f)$ is convex. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$. So, $(\mathbf{x}_1, f(\mathbf{x}_1)) \in epi(f)$ and $(\mathbf{x}_2, f(\mathbf{x}_2)) \in epi(f)$. Since $epi(f)$ is convex, for $\theta \in (0, 1)$,

$$(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2, \theta\alpha_1 + (1-\theta)\alpha_2) \in epi(f)$$

which implies that $f(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq \theta f(\mathbf{x}_1) + (1-\theta)f(\mathbf{x}_2))$ for any $\theta \in (0, 1)$. This proves the sufficiency. $\square$