ML and Opt Lecture 8

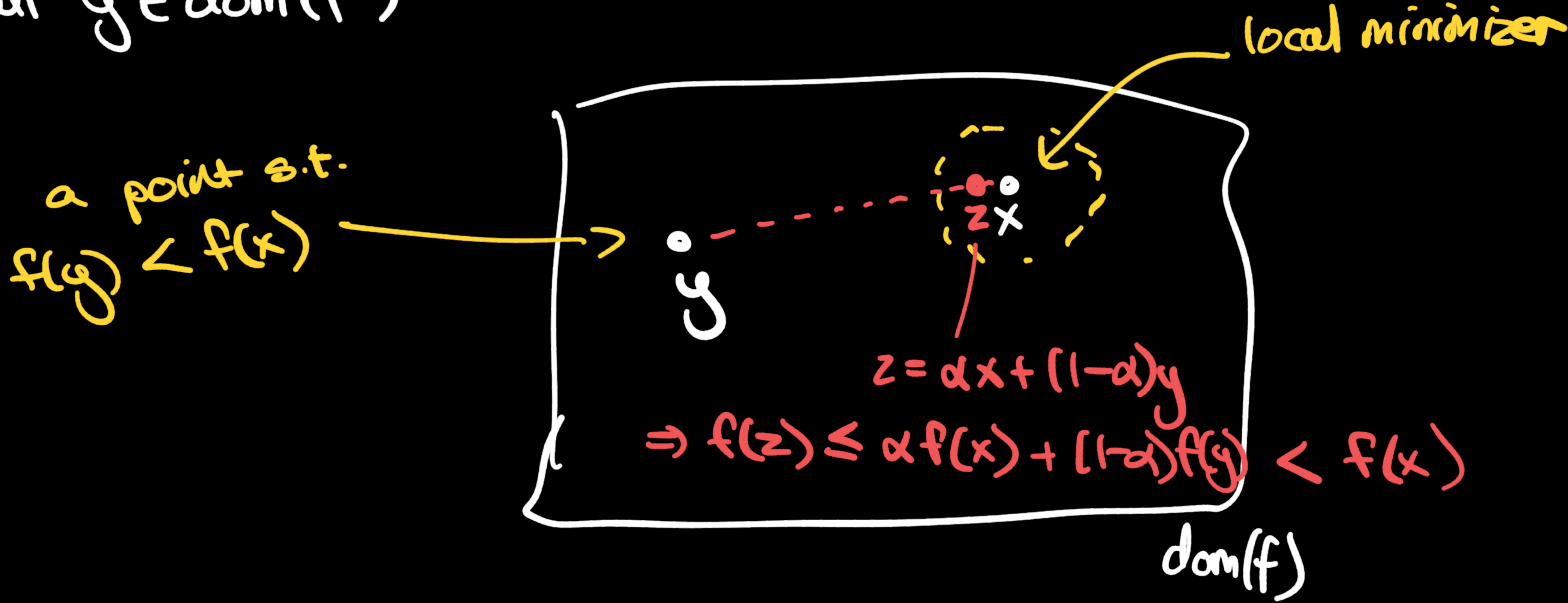
- Most important consequence of convexity:
  local minima are global minima
- Strict convexity & strong convexity:
  strict convexity $\Rightarrow$ a unique minimizer
- First and second-order characterizations of convex functions
- Examples of convex functions using
- Operations that preserve convexity of functions
- Examples of convex optimization problems

Local minima of convex functions are global minima

A point x is a local minimizer of a function f if there exists an $\varepsilon > 0$ s.t. when

$$\|y - x\|_2 < \varepsilon, \quad f(y) \geq f(x)$$

A point x is a global minimizer of f if $f(y) \geq f(x)$ for all $y \in \mathrm{dom}(f)$



Prf

Assume x is a local minimizer but not a global minimizer. Then there exists an $\varepsilon > 0$ so that $f(x) \leq f(z)$ whenever $\|x - z\|_2 < \varepsilon$, and there exists a y s.t. $f(y) < f(x)$.

Consider the line segment $[x, y] = \{\alpha x + (1-\alpha) y : \alpha \in [0,1]\}$

Let $z = \alpha x + (1-\alpha) y$, then $\|x - z\|_2 = \|(1-\alpha)x - (1-\alpha)y\|_2$

$$= (1-\alpha)\|x - y\|_2$$

so by choosing α so that $0 < (1-\alpha)\|x-y\|_2 < \varepsilon$,

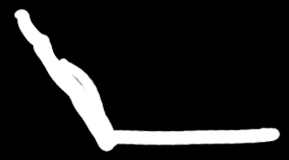
we have z s.t. $\|z - x\|_2 < \varepsilon$ and $z \neq x$

Then we also have

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

by convexity   use $f(y) < f(x)$

This contradicts the local minimality of x. Thus x must be a global minimizer.



$(1-t)_+$ not strictly cvx

Fact

If f is a convex functions, its set of minimizers is a convex set.

x^2

Def Strict convexity

A function f is strictly convex if $\forall x, y \in \mathrm{dom}(f)$

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

x^2

Def Strong convexity

A function f is strongly convex if $\forall x, y \in \mathrm{dom}(f)$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y) + \frac{\alpha(1-\alpha)}{2} \|x-y\|_2^2$$

Facts

- strong convexity $\Rightarrow$ strict convexity $\Rightarrow$ convexity

- If f is strictly convex, then it has a unique global minimizer (if it has a minimizer)

- If f is convex and g is strictly convex (strongly cvx), then

  $f + g$ is strictly convex (strongly cvx)

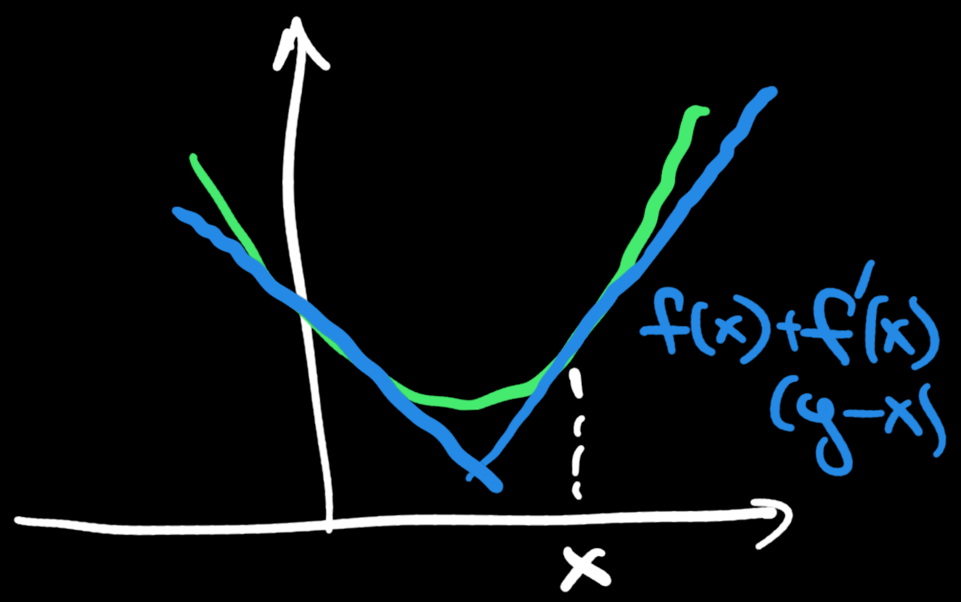
First-order characterizations of convexity

Assume f is a differentiable function

If $f: \mathbb{R} \to \mathbb{R}$, $\forall x, y$

$$f(y) \geq f(x) + f'(x)(y-x)$$

if and only if f is convex



A convex function always lies on or above its first-order TS

If $f: \mathbb{R}^d \to \mathbb{R}$, $\forall x, y$

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

iff f is convex

Ex:

Establishing that x^2 is convex:

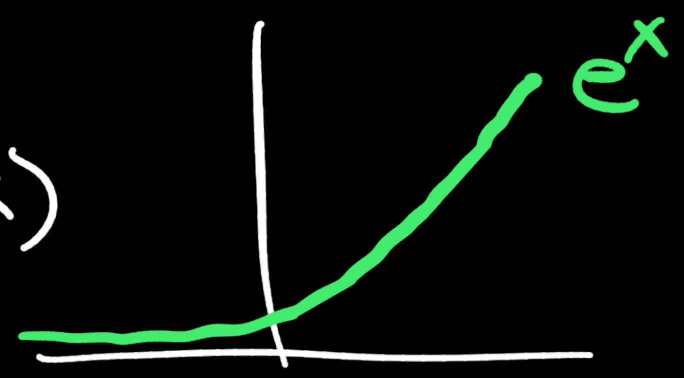
x^2 is convex $\iff \forall x,y: \quad y^2 \geq x^2 + 2x(y-x)$

$\iff \forall x,y: \quad y^2 \geq -x^2 + 2xy$

$\iff \forall x,y: \quad \underbrace{y^2 + x^2 - 2xy}_{=(x-y)^2} \geq 0$

$(x-y)^2 \geq 0$

Establishing that e^x is convex on $\mathbb{R}$:



e^x is convex $\iff \forall x,y: \; e^y \geq e^x + e^x(y-x)$

$\iff \forall x,y: \; e^{y-x} \geq 1 + (y-x)$

Claim: $e^z \geq 1 + z$ when $z > 0$

$e^{-z} \geq 1 - z$ when $z \leq 0$

$\Rightarrow e^z \geq 1+z$ for all $z \in \mathbb{R}$

$\Rightarrow e^x$ is convex

Exercises:

Use first-order characterization to show $\frac{1}{x}$ and $-\log(x)$ are convex

Second-order characterization of convexity

Assume f is second-differentiable and $\mathrm{dom}(f)$ is convex

If $f: \mathbb{R} \to \mathbb{R}$,

$$f''(x) \geq 0 \quad \text{on } \mathrm{dom}(f)$$

iff f is convex

A function is convex iff it's curved upward

If $f: \mathbb{R}^d \to \mathbb{R}$,

$$\nabla^2 f(x) \succcurlyeq 0 \quad \text{on } \mathrm{dom}(f)$$

iff f is convex

Ex:

$f(x) = \frac{1}{x}$ is convex $\Leftrightarrow$ $f''(x) = \frac{2}{x^3} \geq 0$
on $\mathbb{R}_{++}$

Ex logsumexp$(\cdot)$ on $\mathbb{R}^d$, $\underset{f(x)\,=}{\text{logsumexp}(x)} = \log\left(\sum_{i=1}^{d} e^{x_i}\right)$

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]_{\substack{i=1\ldots d \\ j=1\ldots d}} = \frac{\partial}{\partial x_i}\left[\nabla f(x)\right]_j$$

$$\left[\nabla f(x)\right]_j = \frac{1}{\sum_{i=1}^{d} e^{x_i}} e^{x_j} = \left[\text{softmax}(x)\right]_j$$

$$\Rightarrow \nabla f(x) = \text{softmax}(x)$$

$$[\nabla^2 f(x)]_{ij} = \frac{\partial}{\partial x_i}[\mathrm{softmax}(x)]_j = \frac{\partial}{\partial x_i}\left[\frac{e^{x_j}}{1^T e^x}\right]$$

$$= \begin{cases} \dfrac{e^{x_i}(1^T e^x) - e^{2x_i}}{(1^T e^x)^2} = \dfrac{e^{x_i}}{1^T e^x} - \dfrac{e^{2x_i}}{(1^T e^x)^2}, & i = j \\[2ex] \dfrac{-e^{x_i + x_j}}{(1^T e^x)^2}, & i \neq j \end{cases}$$

$$\Rightarrow \nabla^2 f(x) = \mathrm{Diag}(\mathrm{softmax}(x)) - \mathrm{softmax}(x)\,\mathrm{softmax}(x)^T$$

$$:= \mathrm{Diag}(\theta) - \theta\theta^T$$

where $\theta = \mathrm{softmax}(x)$

Recall that $\nabla^2 f(x)$ is PSD if $\forall z \in \mathbb{R}^d$:

$$z^T \nabla^2 f(x) z \geq 0$$

Notice

$$z^T \nabla^2 f(x) z = z^T \left[\mathrm{Diag}(\theta) - \theta\theta^T \right] z$$
$$= \sum_{i=1}^{d} \theta_i z_i^2 - \left(\sum_{i=1}^{d} \theta_i z_i \right)^2$$

Recall that θ is a probability vector $\theta = \mathrm{softmax}(x)$

Let φ be a r.v. s.t. $\mathbb{P}(\varphi = z_i) = \theta_i$, then

$$\mathrm{Var}(\varphi) = \mathbb{E}(\varphi - \mathbb{E}\varphi)^2 \geq 0$$

and

$$\mathrm{Var}(\varphi) = \mathbb{E}[\varphi^2] - (\mathbb{E}\varphi)^2 = \sum_{i=1}^{d} \theta_i z_i^2 - \left(\sum_{i=1}^{d} \theta_i z_i \right)^2$$

so $z^T \nabla^2 f(x) z = \mathrm{Var}(\varphi) \geq 0$ for $\underline{\underline{\text{all}}}$ z $\Rightarrow$ $\underset{\text{cvx}}{f(x)}$

Operations that preserve convexity of functions

- $\|x\|_2$ is convex on $\mathbb{R}^d$:

$$\|\alpha x + (1-\alpha)y\|_2 \overset{\triangle\text{-ineq}}{\leq} \|\alpha x\|_2 + \|(1-\alpha)y\|_2$$

$$\overset{\text{1-homogeneity}}{\leq} \alpha\|x\|_2 + (1-\alpha)\|y\|_2$$

In fact, <u>any</u> norm is convex

- $f(x) = a^T x + b$ (affine functions) are convex

- $f(x) = \frac{1}{2} x^T A x + b^T x + c$

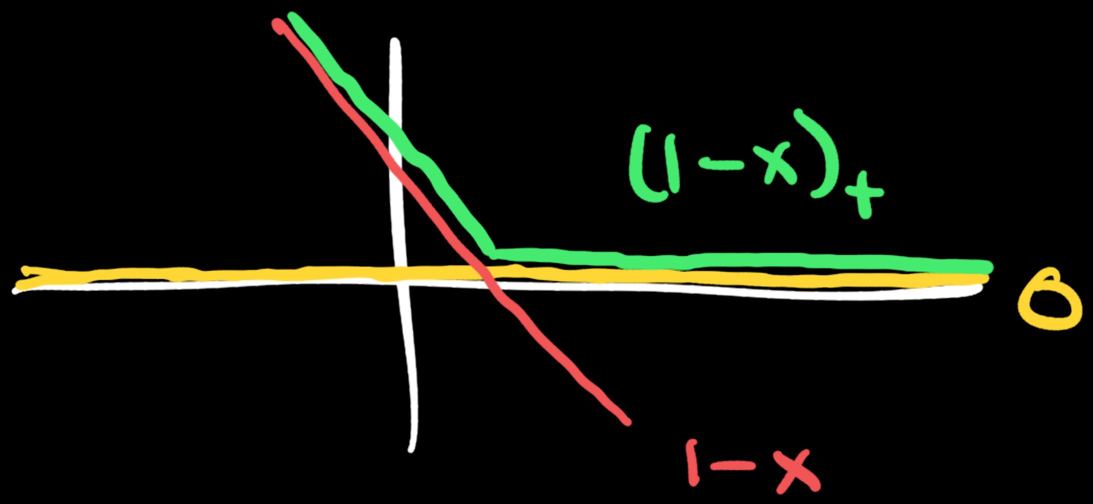
$\nabla^2 f(x) = A$ so f is convex iff $A \succeq 0$

- If $f_1, \dots, f_n$ are convex on a set C and $\alpha_1, \dots, \alpha_n$ are nonnegative scalars, then $\sum_{i=1}^{n} \alpha_i f_i$ is convex on C

- If f is convex on $\mathrm{dom}(f)$, then $f(Ax+b)$ is convex on the set of x s.t. $Ax+b \in \mathrm{dom}(f)$

- The pointwise maximum of convex functions is convex

$$(1-x)_+ = \max(1-x, 0)$$



- If f is a convex and non-decreasing function ($x \geq y$ then $f(x) \geq f(y)$) and g is a convex function, then $f(g(x))$ is a convex function.

<u>Ex</u> $f(x) = x^2$ is convex and non-decreasing on $\mathbb{R}_+$, so if g is convex and nonnegative, then g^2 is convex.

<u>Cor</u> The square of <u>any</u> norm is convex

Examples of convex optim probs

1) Linear programming

$$\min \; c^T x$$
$$Ax \leq b$$

linear functions

intersection of half spaces

2) Least squares

$$\min_x \; \|Ax - b\|_2^2$$

norm squared is cvx, composition of affine w/ cvx is cvx

3) Non-negative LS

$$\min_{x \geq 0} \; \|Ax - b\|_2^2$$

positive orthan is cvx

4) ℓ_2/ℓ_1 - regularized least squares

norms are cvx

$$\underbrace{\min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2 + \lambda \|x\|_{1 \text{ or } 2}}_{\text{positive combination of cvx functions is cvx}}$$

positive regularization