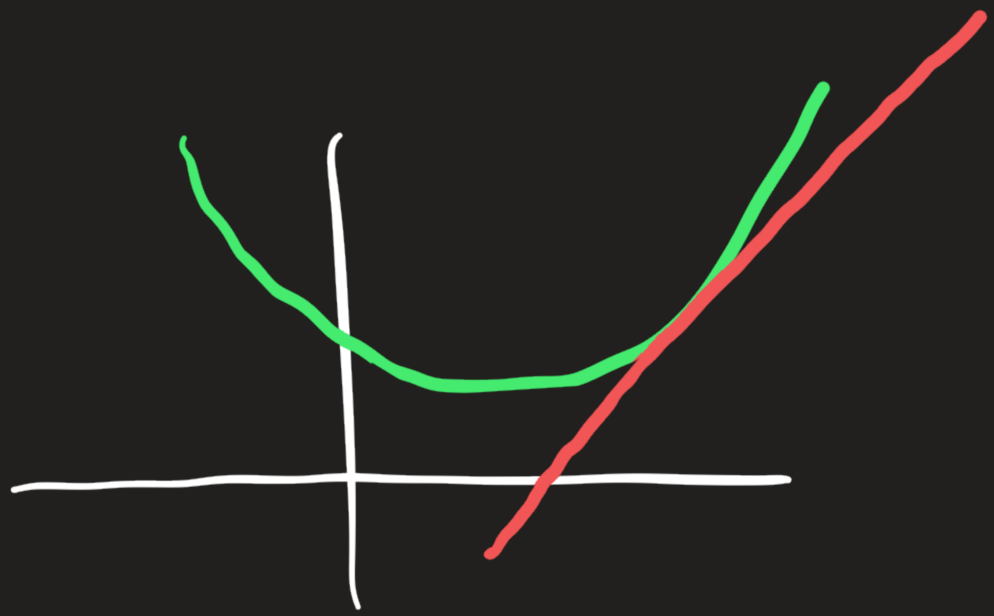
ML & Opt Lecture 7

- First & 2^{nd} order characterizations of convexity
- Examples of convex functions using
- Operations that preserve convexity
- Examples of convex optim problems
- Jensen's inequality

First-order characterization of convexity

A differentiable function whose domain is a convex set is convex iff it lies above its tangent at each point in its domain, i.e.

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) \quad \text{for all } x, y \in \mathrm{dom}(f)$$



Ex: x^2 is convex

1st: $\mathrm{dom}(x \mapsto x^2) = \mathbb{R}$ so the function is defined on a convex set

2nd: By first order characterization

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

x^2 lies above its tangent line at any point x

$$\Leftrightarrow y^2 \geq x^2 - 2x^2 + 2xy$$

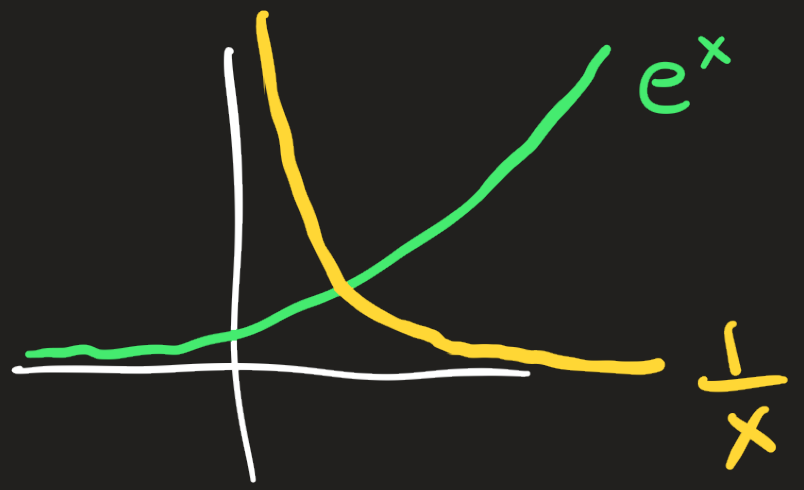
$$\Leftrightarrow y^2 \geq 2xy - x^2$$

$$\Leftrightarrow x^2 + y^2 - 2xy \geq 0$$

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

$\Rightarrow x^2$ is convex

Ex: e^x and



$$\Leftrightarrow e^y \geqslant e^x + e^x(y-x)$$

$$\Leftrightarrow e^y - e^x \geqslant e^x(y-x)$$

$$\Leftrightarrow e^{y-x} - 1 \geqslant y-x$$

$$\Leftrightarrow e^{y-x} \geqslant 1 + (y-x) \quad \checkmark$$

Second-order characterization of convexity (scalar functions)

If f is second-differentiable and dom(f) is convex, then f is convex iff $f''(x) \geq 0$ on dom(f)

$f(x) = \frac{1}{x}$ on $\mathbb{R}_{++}$ then $f''(x) = \left(-\frac{1}{x^2}\right)' = \frac{2}{x^3} > 0$

Second-order characterization of convexity

A twice differentiable function f defined on a convex set is convex iff at each point in $\mathrm{dom}(f)$,

$\nabla^2 f(x) \succcurlyeq 0$ (the Hessian is positive semidefinite (PSD))

Characterizations of PSD:

1) for all $y \in \mathbb{R}^d$ $\langle y, \nabla^2 f(x) y \rangle \geq 0$
2) $\nabla^2 f(x) = LL^T = C^2$
3) $\nabla^2 f(x)$ has all nonnegative eigenvalues

f curves above its tangent plane

Ex: logsumexp$(\cdot)$ on $\mathbb{R}^d$

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

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

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

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

$$[\nabla^2 f(x)]_{ij} = \frac{\partial}{\partial x_i}[\text{softmax}(x)]_j = \frac{\partial}{\partial x_i}\left[\frac{e^{x_j}}{\sum_{k=1}^{d} e^{x_k}}\right]$$

$$= \begin{cases} \frac{\partial}{\partial x_i}\left[\frac{e^{x_i}}{\sum_{k=1}^{d} e^{x_k}}\right] & i = j \\ \frac{\partial}{\partial x_i}\left[\frac{e^{x_j}}{\sum_{k=1}^{d} e^{x_k}}\right] & i \neq j \end{cases}$$

$$= \begin{cases} \frac{e^{x_i}\left(\sum_{k=1}^{d} e^{x_k}\right) - \left(e^{x_i}\right)^2}{\left(\sum_{k=1}^{d} e^{x_k}\right)^2} & i = j \\ \frac{-e^{x_i} e^{x_j}}{\left(\sum_{k=1}^{d} e^{x_k}\right)^2} & i \neq j \end{cases}$$

$$\left[\nabla^2 f(x)\right]_{ij} = \mathrm{Diag}\left(\frac{e^x}{1^T e^x}\right) - \left[\frac{e^x}{1^T e^x}\right]\left[\frac{e^x}{1^T e^x}\right]^T$$

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

where $\theta = \dfrac{e^x}{1^T e^x} = \mathrm{softmax}(x)$

Need to show $\nabla^2 f(x) \succcurlyeq 0$, i.e.

$\forall z$:

$$z^T\left(\mathrm{Diag}(\theta) - \theta\theta^T\right)z \geqslant 0$$

$$\iff \sum_{i=1}^{d} \theta_i z_i^2 - \left(\sum \theta_i z_i\right)^2 \geqslant 0$$

Note that since $\Theta = \text{softmax}(x)$ is a probability vector, we can take

$$Y = z_i \text{ w.p. } \Theta_i$$

Then

$$\text{Var}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 \geq 0$$

and plug in what these are

$$\mathbb{E}(Y^2) = \sum_{i=1}^{d} \Theta_i z_i^2 \quad \text{and} \quad \mathbb{E}Y = \sum_{i=1}^{d} \Theta_i z_i$$

and we see that

$$z^T \left(\text{Diag}(\Theta) - \Theta\Theta^T \right) z = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 = \text{Var}(Y) \geq 0$$

$$\Rightarrow \nabla^2 f(x) = \text{Diag}(\Theta) - \Theta\Theta^T \succeq 0$$

<u>Ex</u> of convex functions & operations preserving convexity

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

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

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

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

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

$$\nabla f(x) = Ax + b$$

$$\nabla^2 f(x) = A$$

so $\nabla^2 f(x) \succcurlyeq 0 \iff f$ is convex

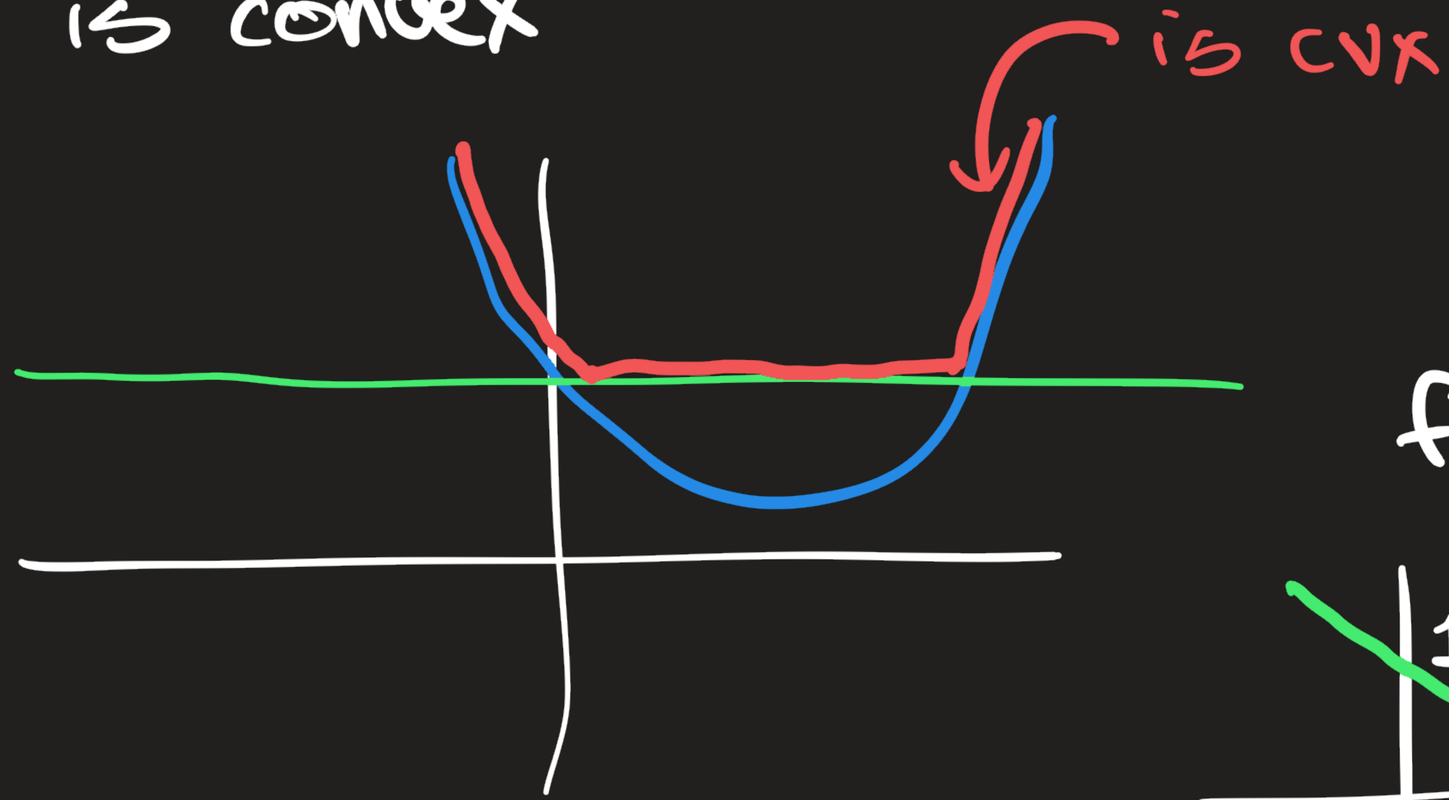
so the quadratic f is convex $\iff A \succcurlyeq 0$

- If $f_1, \dots, f_n$ are convex on some set C and $\alpha_1, \dots, \alpha_n \geq 0$ then

$f = \sum \alpha_i f_i$ is convex on C

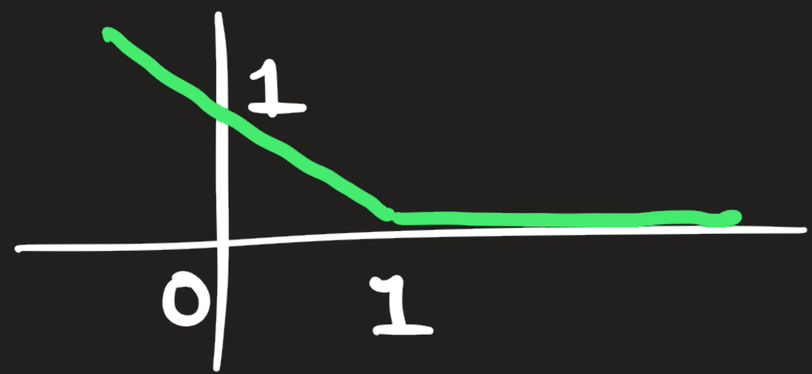
- If f is convex on $\mathrm{dom}(f)$ then $f(Ax+b)$ is convex on the set of x s.t. $Ax+b \in \mathrm{dom}(f)$

- The pointwise maximum of convex functions is convex



Implies:

$$f(t) = (1-t)_+$$



- If f is a convex & non-decreasing function ($x \geq y$ then $f(x) \geq f(y)$) and g is a convex function, then $f \circ g$, defined pointwise as

$$(f \circ g)(x) = f(g(x)),$$

is convex

Ex: $f(x) = x^2$ is nondecreasing & convex on $\mathbb{R}_+$
so if g is convex into $\mathbb{R}_+$, then g^2 is convex

Cor square of any norm is convex

Cor square of any nonnegative cvx function is cvx

Examples of convex optim probs

1) Linear programming

$$\min \; c^T x \quad \text{s.t.} \quad Ax \le b$$

linear function

intersection of half spaces

2) Least squares

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2$$

cvx

norm squared is cvx; cvx of affine is cvx

3) Nonnegative LS

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

positive orthant is cvx

4) ℓ_2/ℓ_1-regularized least squares

$$\min_{x \in \mathbb{R}^d} \|Ax - b\|_2^2 + \lambda \|x\|_{1/2}$$

cvx

positive linear combinations of cvx sets are cvx

5) Logistic regression

$$\min_{\beta \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \log(1 + \exp(-y_i \beta^T x_i))$$

$$= \text{logsumexp}\left(\begin{bmatrix} 0 \\ -y_i \beta^T x_i \end{bmatrix}\right)$$

$$= \text{logsumexp}(M_i \beta) \quad \text{where } M_i = \begin{bmatrix} 0^T \\ -y_i x_i^T \end{bmatrix}$$

Jensen's Inequality

Generalization of the fact that $\mathrm{Var}(X) \geq 0$ for any r.v. X:

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$$
$$= \mathbb{E}(X^2) - (\mathbb{E}X)^2$$
$$= \mathbb{E}f(X) - f(\mathbb{E}X)$$

for $f = x^2$

<u>Jensen's Ineq:</u> For any convex function f and r.v. X taking values in $\mathrm{dom}(f)$:

$$\mathbb{E}f(X) \geq f(\mathbb{E}X)$$

Intuition

Consider X takes on values

$$X = \begin{cases} a & \text{w.p. } p_1 \\ b & \text{w.p. } p_2 \end{cases}$$

by convexity

$$f(p_1 a + p_2 b) \leq p_1 f(a) + p_2 f(b)$$

$$\shortparallel$$

$$f(\mathbb{E}X) \leq \mathbb{E}f(X)$$

Argument for a general r.v.

Assume f is differentiable and convex, then

$$f(X) \geq f(Y) + \nabla f(Y)^T (X - Y)$$

Take $Y = \mathbb{E}X$

$$f(X) \geq f(\mathbb{E}X) + \nabla f(\mathbb{E}X)^T (X - \mathbb{E}X)$$

Take expectations of both sides

$$\mathbb{E}f(X) \geq f(\mathbb{E}X) + \nabla f(\mathbb{E}X)^T (\mathbb{E}X - \mathbb{E}X)$$

$$= f(\mathbb{E}X)$$