

## Lecture 6 – Gradient Descent in Action



DSC 40A, Spring 2023

# Announcements

- ▶ Homework 2 is due **Tuesday at 11:59pm**.
  - ▶ Come to office hours before then for help!
  - ▶ See [dsc40a.com/calendar](https://dsc40a.com/calendar) for the office hours schedule.
- ▶ Solutions to Groupwork 2 and Homework 1 are available on Campuswire.
  - ▶ Reviewing them will help you write better solutions in future assignments.
  - ▶ You should also make sure you know how to do all groupwork and homework questions for exams.

# Agenda

- ▶ Brief recap of Lecture 5.
- ▶ Gradient descent demo.
- ▶ When is gradient descent guaranteed to work?

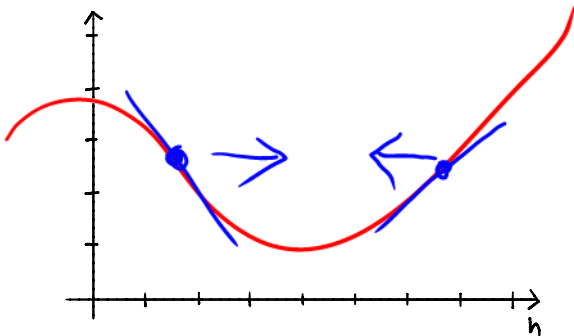
# Gradient descent fundamentals

## The general problem

- ▶ **Given:** a differentiable function  $R(h)$ .
- ▶ **Goal:** find the input  $h^*$  that minimizes  $R(h)$ .

## Key idea behind **gradient descent**

- ▶ If the slope of  $R$  at  $h$  is **positive** then we'll **decrease**  $h$ .
- ▶ If the slope of  $R$  at  $h$  is **negative** then we'll **increase  $h$ .**



## Gradient descent

- ▶ Pick a positive constant,  $\alpha$ , for the **learning rate**.
- ▶ Pick a starting prediction,  $h_0$ .
- ▶ Repeatedly apply the gradient descent update rule.

$$h_i = h_{i-1} - \alpha \cdot \frac{dR}{dh}(h_{i-1})$$

old pred minus a constant  
multiple  
of slope

- ▶ Repeat until convergence (when  $h$  doesn't change much).

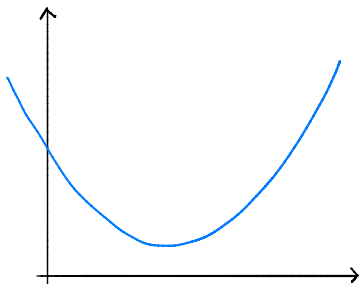
## Gradient descent demo



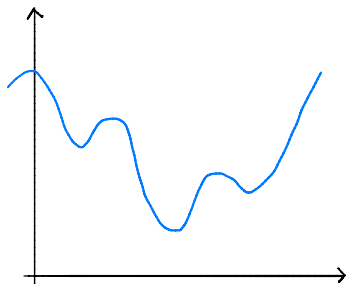
Let's see gradient descent in action. [Follow along here.](#)

**When is gradient descent guaranteed to work?**

# Convex functions



**Convex**



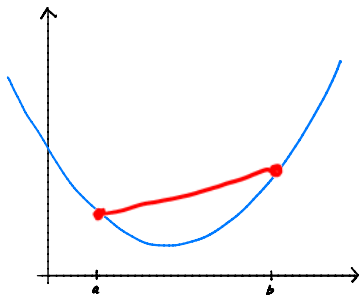
**Non-convex**

## Convexity: Definition

- ▶  $f$  is **convex** if for **every**  $a, b$  in the domain of  $f$ , the line segment between

$$(a, f(a)) \quad \text{and} \quad (b, f(b))$$

does not go below the plot of  $f$ .

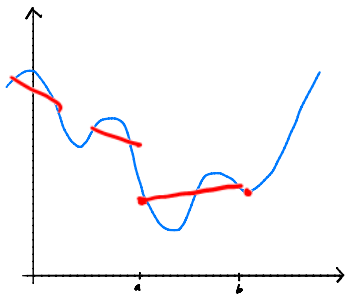


## Convexity: Definition

- ▶  $f$  is **convex** if for **every**  $a, b$  in the domain of  $f$ , the line segment between

$$(a, f(a)) \quad \text{and} \quad (b, f(b))$$

does not go below the plot of  $f$ .



## Convexity: Formal definition

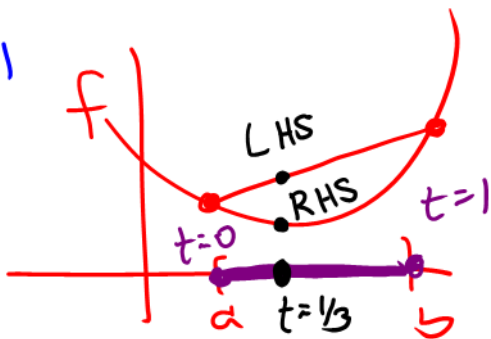
LHS  $\triangleright$  A function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is **convex** if for every choice of  $a, b$  and  $t \in [0, 1]$ :

parameterized line  $\rightarrow$

$$\underbrace{(1-t)f(a) + tf(b)}_{\text{line}} \geq \underbrace{f((1-t)a + tb)}_{\text{function}}$$

$\triangleright$  This is a formal way of restating the condition from the previous slide.

from  $f(a)$  when  $t=0$   
to  $f(b)$  when  $t=1$



## Discussion Question

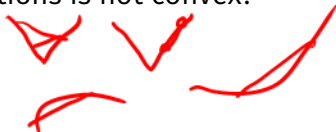
Which of these functions is not convex?

~~a)  $f(x) = |x|$~~

~~b)  $f(x) = e^x$~~

**c)  $f(x) = \sqrt{x - 1}$**

d)  $f(x) = (x - 3)^{24}$



## Why does convexity matter?

- ▶ Convex functions are (relatively) easy to minimize with gradient descent.
- ▶ **Theorem:** if  $R(h)$  is convex and differentiable then gradient descent converges to a **global minimum** of  $R$  *provided* that the step size is small enough.
- ▶ **Why?**
  - ▶ If a function is convex and has a local minimum, that local minimum must be a global minimum.
  - ▶ In other words, gradient descent won't get stuck/terminate in local minimums that aren't global minimums.



## Nonconvexity and gradient descent

- ▶ We say a function is **nonconvex** if it does not meet the criteria for convexity.
- ▶ Nonconvex functions are (relatively) hard to minimize.
- ▶ Gradient descent can still be useful, but it's not guaranteed to converge to a global minimum.
  - ▶ We saw this when trying to minimize  $R_{ucsd}(h)$  with a smaller  $\sigma$ .

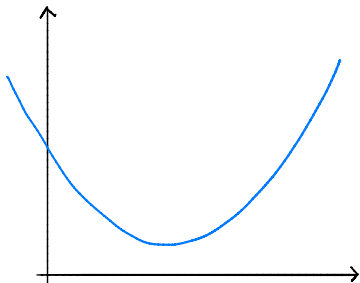
## Second derivative test for convexity

- ▶ If  $f(x)$  is a function of a single variable and is twice differentiable, then:
- ▶  $f(x)$  is convex if and only if  $\frac{d^2f}{dx^2}(x) \geq 0$  for all  $x$ .

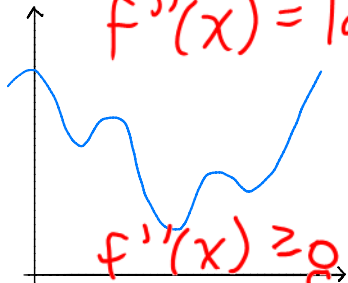
▶ Example:  $f(x) = x^4$  is convex.

$\Rightarrow f'(x) = 4x^3$   
 $f''(x) = 12x^2$

✓  
 $\geq 0$   
for  
all  
 $x$   
for  
 $x^2$



Convex



Non-convex

## Convexity of empirical risk

- ▶ If  $L(h, y)$  is a convex function (when  $y$  is fixed) then

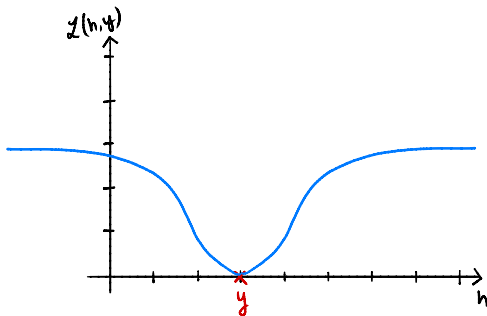
$$R(h) = \frac{1}{n} \sum_{i=1}^n L(h, y_i)$$

is convex.

- ▶ More generally, sums of convex functions are convex.
- ▶ What does this mean?
  - ▶ If a loss function is convex, then the corresponding empirical risk will also be convex.

# Convexity of loss functions

- ▶ Is  $L_{\text{sq}}(h, y) = (y - h)^2$  convex? **Yes** or **No**.
- ▶ Is  $L_{\text{abs}}(h, y) = |y - h|$  convex? **Yes** or **No**.
- ▶ Is  $L_{\text{ucsd}}(h, y)$  convex? **Yes** or **No**.



## Convexity of $R_{ucsd}$

- ▶ A function can be convex in a region.
- ▶ If  $\sigma$  is large,  $R_{ucsd}(h)$  is convex in a big region around data.
  - ▶ A large  $\sigma$  led to a very smooth, parabolic-looking empirical risk function with a single local minimum (which was a global minimum).
- ▶ If  $\sigma$  is small,  $R_{ucsd}(h)$  is convex in only small regions.
  - ▶ A small  $\sigma$  led to a very bumpy empirical risk function with many local minimums.

## Discussion Question

Recall the empirical risk for absolute loss,

$$R_{abs}(h) = \frac{1}{n} \sum_{i=1}^n |y_i - h|$$

Is  $R_{abs}(h)$  **convex**? Is gradient descent **guaranteed** to find a global minimum, given an appropriate step size?

- a) **YES** convex, **YES** guaranteed
- b) **YES** convex, **NOT** guaranteed
- c) **NOT** convex, **YES** guaranteed
- c) **NOT** convex, **NOT** guaranteed

## Summary

## Summary

- ▶ Gradient descent is a general tool used to minimize differentiable functions.
- ▶ Convex functions are (relatively) **easy** to optimize with gradient descent.
- ▶ We like **convex loss functions**, such as the squared loss and absolute loss, because the corresponding empirical risk functions are also convex.



## What's next?

- ▶ So far, we've been predicting future values (salary, for instance) without using any information about the individual.
  - ▶ GPA.
  - ▶ Years of experience.
  - ▶ Number of LinkedIn connections.
  - ▶ Major.
- ▶ How do we incorporate this information into our prediction-making process?