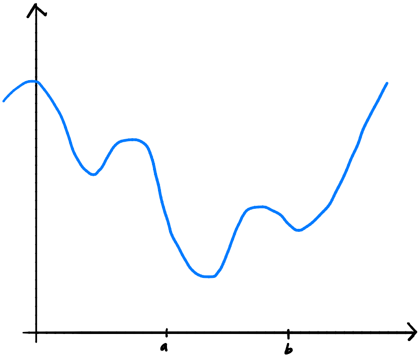


CSE 151A

Intro to Machine Learning

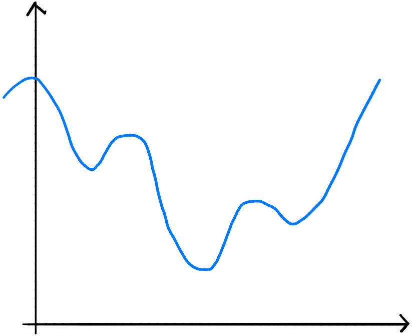
Lecture 10 – Part 01

Convexity in 1-d

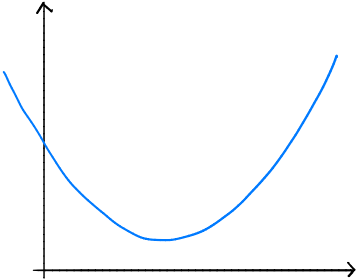
Today

When is gradient descent guaranteed to work?

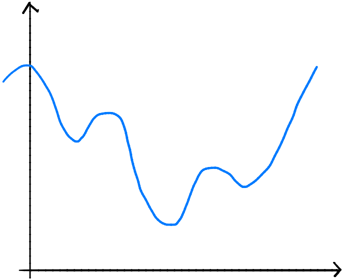
Not here...



Convex Functions



Convex



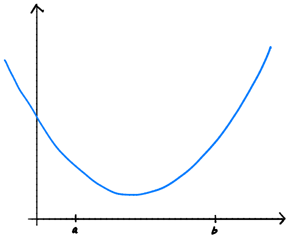
Non-convex

Convexity: Definition

- f is **convex** if for **every** a, b the line segment between

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

does not go below the plot of f.

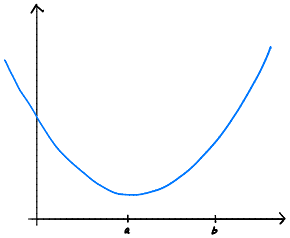


Convexity: Definition

- f is **convex** if for **every** a, b the line segment between

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

does not go below the plot of f.

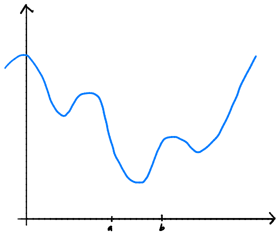


Convexity: Definition

- f is **convex** if for **every** a, b the line segment between

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

does not go below the plot of f.

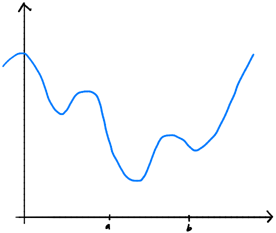


Convexity: Definition

- f is **convex** if for **every** a, b the line segment between

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

does not go below the plot of f.



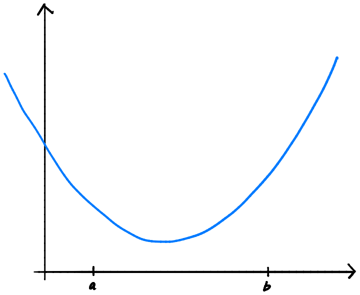
Other Terms

- If a function is not convex, it is **non-convex**.
- **Strictly convex**: the line lies strictly above curve.
- **Concave**: the line lines on or below curve.

Convexity: Formal Definition

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

$$(1-t)f(a) + tf(b) \geq f((1-t)a + tb).$$

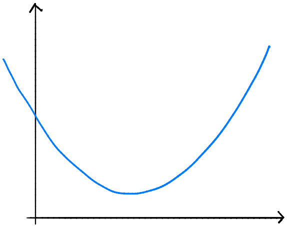


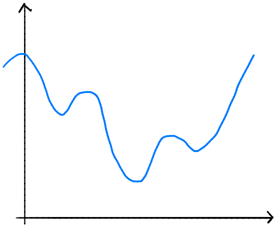
Example

Is $f(x) = |x|$ convex?

Another View: Second Derivatives

- If $\frac{d^2 f}{dx^2}(x) \geq 0$ for all x, then f is convex.
- Example: $f(x) = x^4$ is convex.
- **Warning!** Only works if f is twice differentiable!





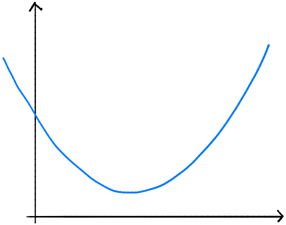
Another View: Second Derivatives

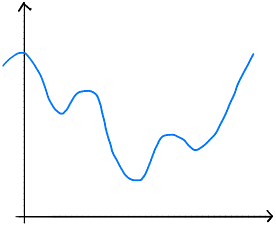
- "Best" straight line at x_0:
  - $h_1(z) = f'(x_0) \cdot z + b$

- "Best" parabola at x_0:
  - At x_0, f looks likes $h_2(z) = \frac{1}{2} f''(x_0) \cdot z^2 + f'(x_0) z + c$
  - Possibilities: upward-facing, downward-facing.

Convexity and Parabolas

- Convex if for **every** x_0, parabola is upward-facing.
  - That is, $f''(x_0) \geq 0$.





Convexity and Gradient Descent

- Convex functions are (relatively) easy to optimize.
- **Theorem**: if $R(x)$ is convex and differentiable[1][2] then gradient descent converges to a **global optimum** of R *provided* that the step size is small enough[3].

---

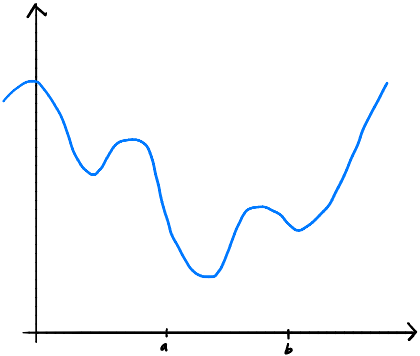
[1] and its derivative is not too wild

[2] actually, a modified GD works on non-differentiable functions

[3] step size related to steepness.

Nonconvexity and Gradient Descent

- Nonconvex functions are (relatively) hard to optimize.
- Gradient descent can still be useful.
- But not guaranteed to converge to a global minimum.



CSE 151A

Intro to Machine Learning

Lecture 10 – Part 02

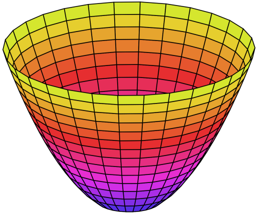
Convexity in Many Dimensions

Convexity: Definition

- $f(\vec{x})$ is **convex** if for **every** $\vec{a}, \vec{b}$ the line segment between

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

  does not go below the plot of f.



Convexity: Formal Definition

- A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if for every choice of $\vec{a}, \vec{b} \in \mathbb{R}^d$ and $t \in [0, 1]$:

$$(1 - t)f(\vec{a}) + tf(\vec{b}) \geq f((1 - t)\vec{a} + t\vec{b}).$$

The Second Derivative Test

- For 1-d functions, convex if second derivative ≥ 0.
- For 2-d functions, convex if ???

Second Derivatives in 2-d

- In 2-d, there are 4 second derivatives of $f(\vec{x})$:
  - $\frac{\partial f^2}{\partial x_1^2}, \frac{\partial f^2}{\partial x_2^2}, \frac{\partial f^2}{\partial x_1 x_2}, \frac{\partial f^2}{\partial x_2 x_1}$

Convexity in 2-d

- "Best" quadratic function approximating f at $\vec{x}$:

$$
\begin{aligned}
h_2(z_1, z_2) &= az_1^2 + bz_2^2 + cz_1 z_2 + \ldots \\
&= \frac{1}{2}\frac{\partial f^2}{\partial x_1^2}(\vec{x}) \cdot z_1 + \frac{1}{2}\frac{\partial f^2}{\partial x_2^2}(\vec{x}) \cdot z_2 + \frac{\partial f^2}{\partial x_1 x_2}(\vec{x}) \cdot z_1 z_2 + \ldots
\end{aligned}
$$

- a, b, c determine rough shape. Possibilities:
  - Upward-facing bowl.
  - Downward-facing bowl.
  - "Saddle"

Convexity in 2-d

- Convex if at any $\vec{x}$, for any z_1, z_2:

$$\frac{1}{2}\frac{\partial f^2}{\partial x_1^2}(\vec{x}) \cdot z_1 + \frac{1}{2}\frac{\partial f^2}{\partial x_2^2}(\vec{x}) \cdot z_2 + \frac{\partial f^2}{\partial x_1 x_2}(\vec{x}) \cdot z_1 z_2 \geq 0$$

The Hessian Matrix

- Create the **Hessian** matrix of second derivatives:

$$H(\vec{x}) = \begin{pmatrix} \frac{\partial f^2}{\partial x_1^2}(\vec{x}) & \frac{\partial f^2}{\partial x_1 x_2}(\vec{x}) \\ \frac{\partial f^2}{\partial x_2 x_1}(\vec{x}) & \frac{\partial f^2}{\partial x_2^2}(\vec{x}) \end{pmatrix}$$

In General

- If $f : \mathbb{R}^d \to \mathbb{R}$, the **Hessian** at $\vec{x}$ is:

$$H(\vec{x}) = \begin{pmatrix} \frac{\partial f^2}{\partial x_1^2}(\vec{x}) & \frac{\partial f^2}{\partial x_1 x_2}(\vec{x}) & \cdots & \frac{\partial f^2}{\partial x_1 x_d}(\vec{x}) \\ \frac{\partial f^2}{\partial x_2 x_1}(\vec{x}) & \frac{\partial f^2}{\partial x_2^2}(\vec{x}) & \cdots & \frac{\partial f^2}{\partial x_2 x_d}(\vec{x}) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f^2}{\partial x_d x_1}(\vec{x}) & \frac{\partial f^2}{\partial x_d^2}(\vec{x}) & \cdots & \frac{\partial f^2}{\partial x_d^2}(\vec{x}) \end{pmatrix}$$

Observations

- H is square.
- H is symmetric.

Convexity in 2-d

- Convex if at any $\vec{x}$, for any z_1, z_2:

$$\frac{1}{2}\frac{\partial f^2}{\partial x_1^2}(\vec{x}) \cdot z_1 + \frac{1}{2}\frac{\partial f^2}{\partial x_2^2}(\vec{x}) \cdot z_2 + \frac{\partial f^2}{\partial x_1 x_2}(\vec{x}) \cdot z_1 z_2 \geq 0$$

- Equivalently, convex if for any $\vec{x}$ and any $\vec{z}$:

$$\vec{z}^T H(\vec{x})\,\vec{z} \geq 0$$

Positive Semi-Definite

- A square, $d \times d$ symmetric matrix X is **positive semi-definite** (PSD) if for any $\vec{u}$:

$$\vec{u}^T X \vec{u} \geq 0$$

The Second Derivative Test

- A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if for any $\vec{x} \in \mathbb{R}^d$, the Hessian matrix $H(\vec{x})$ is positive semi-definite.

But wait...

- How can we tell if a matrix is positive semi-definite?

Example

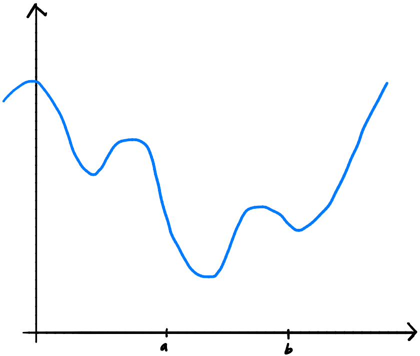
$$M = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Example

$$M = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

Example

Is $f(x, y) = x^2 + 4xy + y^2$ convex?



CSE 151A

Intro to Machine Learning

Lecture 10 – Part 03

Convex Loss Functions

Convexity and Gradient Descent

- Convex functions are (relatively) easy to optimize.

- **Theorem**: if $R(\vec{w})$ is convex and differentiable[4][5] then gradient descent converges to a **global optimum** of R *provided* that the step size is small enough[6].

---

[4] and its gradient is not too wild

[5] actually, a modified GD works on non-differentiable functions

[6] step size related to steepness.

Example

- Recall the Mean Squared Error:

$$R(\vec{w}) = \frac{1}{n} \sum_{i=1}^{n} \left(\vec{x}^{(i)} \cdot \vec{w} - y_i \right)^2$$

- Is this convex?

Mean Squared Error

$$R(\vec{w}) = \frac{1}{n} \sum_{i=1}^{n} \left(\vec{x}^{(i)} \cdot \vec{w} - y_i \right)^2$$

A Useful Theorem

- A square, symmetric matrix M is PSD if and only if M can be written as UU^T for some matrix U.

$$
\begin{aligned}
\vec{x}^T M \vec{x} &= \vec{x}^T U U^T \vec{x} \\
&= (\vec{x}^T U)(U^T \vec{x}) \\
&= (U^T \vec{x})^T (U^T \vec{x}) \\
&= \|U^T \vec{x}\|^2 \\
&\geq 0
\end{aligned}
$$

Mean Squared Error

- The MSE is a convex function of $\vec{w}$.
- We had an explicit solution for the best $\vec{w}$:

$$\vec{w} = (X^T X)^{-1} X^T \vec{y}$$

- But we could also have used gradient descent.

Logistic Regression

- The log-likelihood is **concave**.

$$\log \mathcal{L}(\vec{w}) = -\sum_{i=1}^{n} \log\left[1 + e^{-y_i \vec{w} \cdot \text{Aug}(\vec{x}^{(i)})}\right]$$

Status Update

- We learned what it means for a function to be **convex**.
- Convex functions are (relatively) **easy** to optimize with gradient descent.
- We like **convex loss functions**, like the mean squared error.