Multivariable Calculus
Optimization

06/03/2024

Optimization: Motivations

> *'Nothing takes place in the world whose meaning is not that of some maximum or minimum.'* Leonhard Euler

`https://www.youtube.com/watch?v=IHZwWFHWa-w`

Motivations

- Economics
- Data analysis
- Machine learning
- Physics
- Even the stuff that seems the most random like flipping a coin (and Brownian motion) comes from an optimization problem!

Some toy problems: Economics

A store tries to optimize its sale of Apple Macbook Air and Lenovo Thinkpad. A Macbook Air has the retail price \$1000 and a Thinkpad \$700. The importation costs for the two laptops would be \$900 for the Macbook Air and \$500 for the Thinkpad. The initial capital of the store is $\$100,000$.

How many Macbook Airs and Thinkpads should the store import if it wants to optimize its profit?

Some toy problems: Data analysis

Suppose a scientist has a collection of paired data $\{(x_i, y_i)\}_{i=1}^N$. What is the formula for line that "best fits" with the data?

Some toy problems: Machine learning

What is the best label for the following?

Some toy problems: Machine learning

What is the best label for the following?

(Cat)

Some toy problems: physics

Fermat's principle: light takes path that minimizes the time it travels.

What is the path?

Some toy problems: information theory

A long string of characters is picked from the 26 alphabets.

How should we distribute the alphabets so that the string looks as chaotic as possible?

Some toy problems: information theory

A long string of characters is picked from the 26 alphabets.

How should we distribute the alphabets so that the string looks as chaotic as possible?

Claude Shannon's proposal: use the entropy function

$$-\sum_{i=1}^{26} p(x_i) \log(p(x_i)),$$

where x_i is the ith alphabet and $p(x_i)$ is the probability that x_i will appear at random. This measures how chaotic a certain probabilistic behavior could be.

https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf

https://www.youtube.com/watch?v=v68zYyaEmEA

Caution

The field of optimization as a whole is a very old field. However, regardless of its long history (over 200 years now, since Newton), there's not much one could do when it comes to real life.

There are a lot of questions still remained open in this field.

The "most approachable" kind of optimization is called *convex optimization*.

Caution

The field of optimization as a whole is a very old field. However, regardless of its long history (over 200 years now, since Newton), there's not much one could do when it comes to real life.

There are a lot of questions still remained open in this field.

The "most approachable" kind of optimization is called *convex optimization*.

People manage to study this without differentiability conditions.

Caution

The field of optimization as a whole is a very old field. However, regardless of its long history (over 200 years now, since Newton), there's not much one could do when it comes to real life.

There are a lot of questions still remained open in this field.

The "most approachable" kind of optimization is called *convex optimization*.

People manage to study this without differentiability conditions.

People also use techniques in convex optimization to study non-convex optimization, which is still an untamed beast.

Let's start with the basics, when everything is nice...

Question: what way do you go when you climb a mountain?

Definition
A function $f : \mathbb{R}^n \to \mathbb{R}$ has a **local maximum** at $\mathbf{x_0}$ if $f(\mathbf{x_0}) \geq f(\mathbf{x})$ for $\mathbf{x} \in B_\delta(\mathbf{x_0})$ for small enough δ. f has a **global maximum** at $\mathbf{x_0}$ if $f(\mathbf{x_0}) \geq f(\mathbf{x})$ for $\mathbf{x} \in D$. f has a **local (global) minimum** at $\mathbf{x_0}$ if $-f$ has a local (global) maximum at $\mathbf{x_0}$

A necessary condition

Theorem (First derivative test)
Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a function. If* $\mathbf{x_0}$ *is a local minimum and* f *has partial derivatives at* $\mathbf{x_0}$*. Then*

$$\partial_{x_i} f(\mathbf{x}_0) = 0 \, .$$

Definition
x_0 is said to be a critical point of $f : \mathbb{R}^n \to \mathbb{R}$ if

$$[Df]_{x_0} = 0$$

or one of the partial derivatives $\partial_{x_i} f(x_0)$ fails to exist.

Second derivative test for 2D

Suppose the second derivatives of $f : \mathbb{R}^2 \to \mathbb{R}$ are continuous near a and suppose a is a critical point. Let

$$D = \partial^2_{x_1} f(a) \partial^2_{x_2} f(a) - (\partial^2_{x_1 x_2} f(a))^2.$$

1. If $D > 0$ and $\partial^2_{x_1} f(a) > 0$ then $f(a)$ is a local minimum.
2. If $D > 0$ and $\partial^2_{x_1} f(a) < 0$ then $f(a)$ is a local maximum.
3. If $D < 0$, then $f(a)$ is neither a local minimum nor local maximum.
4. If $D = 0$, then we cannot conclude.

D is called the discriminant of the function f at a.