Optimization for Machine Learning – CSED490Y

Week 01-2: Introduction

Namhoon Lee

POSTECH

Spring 2022

a quick recap of the course logistics

Optimization everywhere

Optimization is used in many decision science and in the analysis of physical systems.

Optimization everywhere

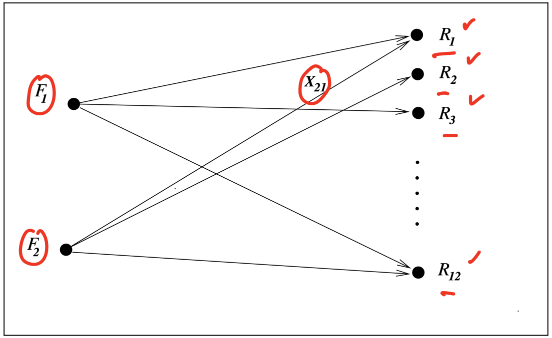
Optimization is used in many decision science and in the analysis of physical systems.

Some examples:

- investment portfoilo for high rate of return
- manufacturing for efficient design and operation of production processes
- circuit design to optimize the performance of electronic devices
- computer program to learn from "experience" with respect to a certain task

Transportation problem

Suppose you want to optimize for a transportation problem.



- There are two factories (F_1, F_2) and a dozen retail outlets (R_1, R_2, .. , R_{12}).
- Requirements: amount of production, demand, cost of shipping, etc.
- Determine how much of the product to ship from each factory to each outlet (x_{ij}) so as to satisfy all the requirements and minimize cost?

Optimization for machine learning

It's about finding settings for some parameters of a system to optimize something.

Optimization for machine learning

$\{0, 1\}$

It's about finding settings for some parameters of a system to optimize something.

	Harry Potter	The Triplets of Belleville	Shrek	The Dark Knight Rises	Memento
	✓		✓	✓	?
		✓			✓
	✓	✓	✓		
				✓	✓

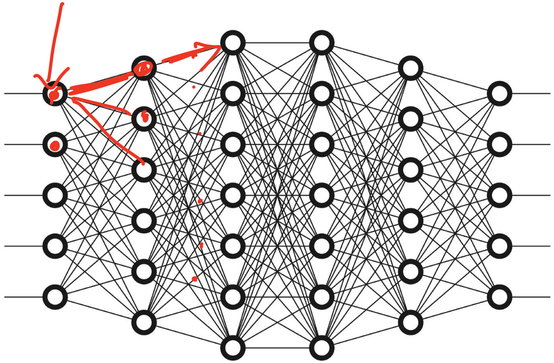
Optimization for machine learning

It's about finding settings for some parameters of a system to optimize something.

Optimization for machine learning

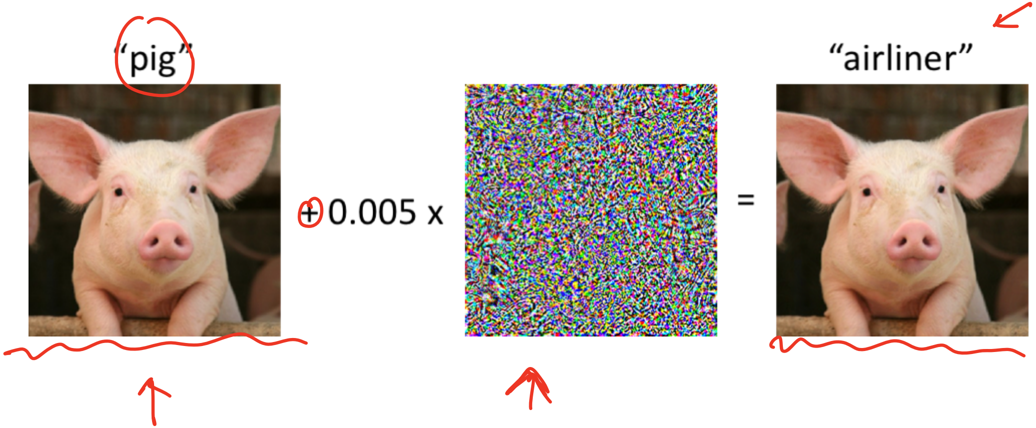
It's about finding settings for some parameters of a system to optimize something.

data



Optimization for machine learning

It's about finding settings for some parameters of a system to optimize something.

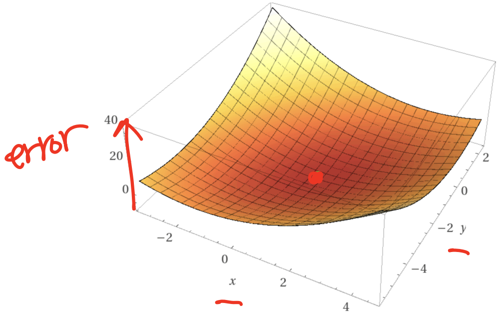


Optimization for machine learning

In machine learning, we often attempt to mimize some cost/error/risk/loss.

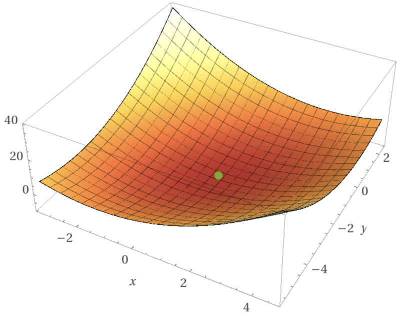
Optimization for machine learning

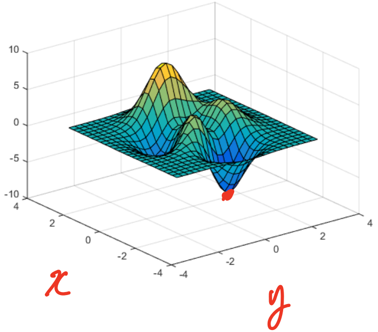
In machine learning, we often attempt to mimize some cost/error/risk/loss.



Optimization for machine learning

In machine learning, we often attempt to mimize some cost/error/risk/loss.

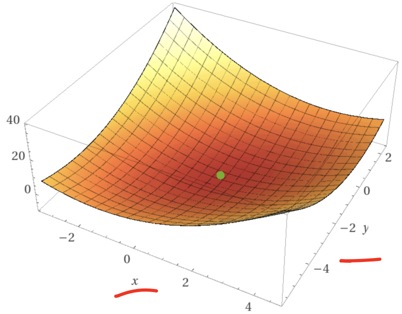


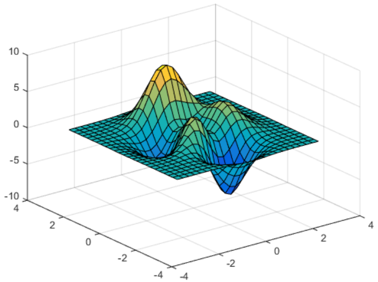


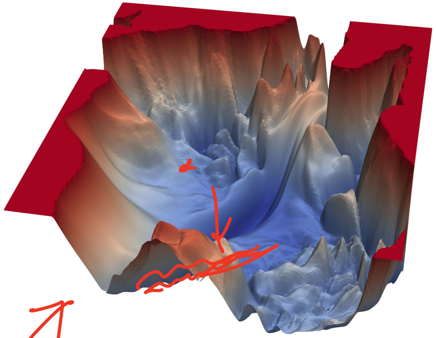
Optimization for machine learning

In machine learning, we often attempt to mimize some cost/error/risk/loss.

f







f(x,y)

Elements of optimization process

Objective

- a quantitative measure of the performance of the system under study
- profit, time, potential energy, or any quantity or combination of quantities that can be represented by a single number

Elements of optimization process

Variables or unknowns / parameters

- certain characteristics of the system that the objective depends on
- find the best possible settings for these variables
- often variables are restricted or constrained

Elements of optimization process

Modeling

- the process of identifying objective, variables, and constraints for a given problem
- construction of an appropriate model is perhaps the most important step
- too simplistic, not give useful insights into the practical problem
- too complex, too difficult to solve

Elements of optimization process

Optimization algorithm

- usually with the help of a computer
- no universal algorithm; rather tailored to a particular type of problem
- the responsibility of choosing which algorithm falls on the user
- determines how fast or slow we can find a solution or whether we can find it at all

Elements of optimization process

Optimality conditions

- to check that the current set of variables is indeed the solution of the problem

Mathematical formulation

An optimization problem:

$$\boxed{\min_{x \in R^n} f(x)}$$
$$\text{s.t. } c_i(x) = 0, \quad i \in \mathcal{E},$$
$$c_i(x) \geq 0, \quad i \in \mathcal{I}.$$

x^*

Mathematical formulation

An optimization problem:

$$\boxed{\begin{aligned} \min_{x\in R^n} \;& f(x) \\ \text{s.t. } & c_i(x) = 0, \;\; i \in \mathcal{E}, \\ & c_i(x) \geq 0, \;\; i \in \mathcal{I}. \end{aligned}}$$

- x: variables, unknowns, parameters
- f: objective function
- c_i: constraint functions
- $\mathcal{E}$ and $\mathcal{I}$: set of indices for equality and inequality constraints

Mathematical formulation

Example:

$$\min\ (x_1 - 2)^2 + (x_2 - 1)^2$$
$$\text{s.t. } x_1^2 - x_2 \le 0,$$
$$x_1 + x_2 \le 2.$$

$$x^* = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$f(x^*) = \underline{0}$$

Mathematical formulation

Example:

$x \in \mathbb{R}^2$

$$\begin{aligned} \min\ & (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{s.t. } & x_1^2 - x_2 \le 0, \\ & x_1 + x_2 \le 2. \end{aligned}$$

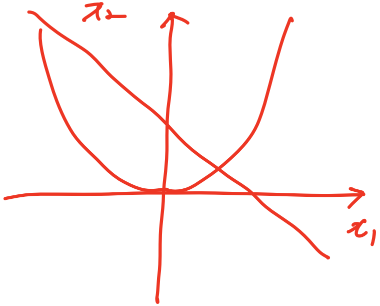
Here, $f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \end{bmatrix} = \begin{bmatrix} -x_1^2 + x_2(x) \\ -x_1 - x_2 + 2(x) \end{bmatrix}$,
$\mathcal{I} = \{1, 2\}$, $\mathcal{E} = \emptyset$.

Mathematical formulation

Example:

2 1

$$\begin{aligned} \min \ & (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{s.t. } & x_1^2 - x_2 \le 0, \\ & x_1 + x_2 \le 2. \end{aligned}$$

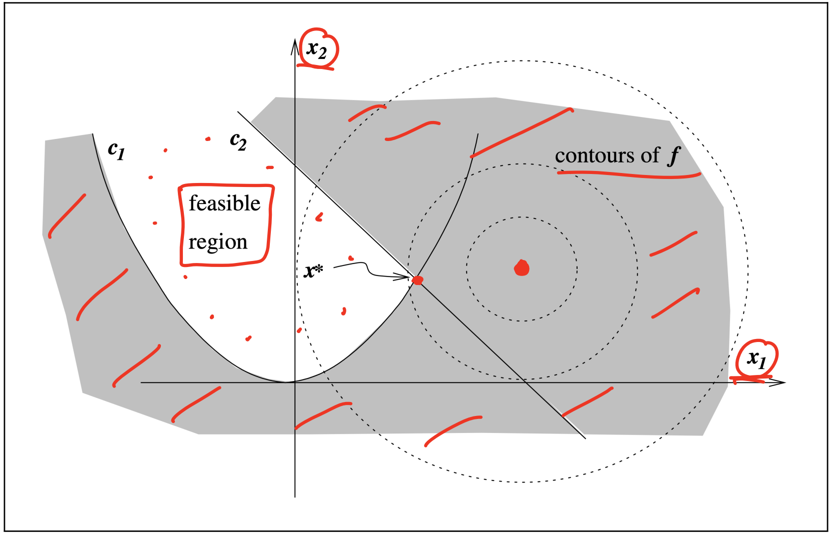


Here, $f(x) = (x_1 - 2)^2 + (x_2 - 1)^2$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \end{bmatrix} = \begin{bmatrix} -x_1^2 + x_2(x) \\ -x_1 - x_2 + 2(x) \end{bmatrix}$, $\mathcal{I} = \{1, 2\}$, $\mathcal{E} = \emptyset$.

Can we illustrate this?

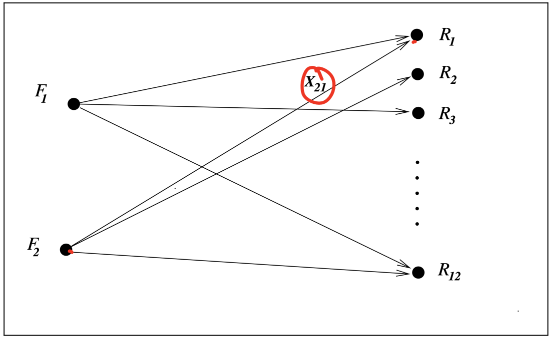
$x^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Mathematical formulation



Transportation problem

Suppose you want to optimize for a transportation problem.



Modeling:

- a_i: amount of product F_i produces each week
- b_j: weekly demand of the product by R_j
- c_{ij}: cost of shipping the product from F_i to R_j
- x_{ij}: amount of product shipped from F_i to R_j

Transportation problem

Writing into a mathematical optimization forumation

$$
\begin{aligned}
\min \quad & \sum_{ij} \boxed{c_{ij} x_{ij}} \\
\text{s.t.} \quad & \sum_{j=1}^{12} x_{ij} \le a_i, \quad i = 1, 2, \\
& \sum_{i=1}^{2} x_{ij} \ge b_j, \quad j = 1, \dots, 12, \\
& x_{ij} \ge 0, \quad i = 1, 2, \; j = 1, \dots, 12.
\end{aligned}
$$

$F_i \xrightarrow{x_{ij}} R_j$

- a.k.a. linear programming
- may turn into non-linear programming with additional conditions

Various forms

Constrained optimization (vs unconstrained)

- When there are constraints on the variables.

Various forms

Constrained optimization (vs unconstrained)

- When there are constraints on the variables.

Discrete optimization (vs continuous)

- When variables only make sense to be discrete values.

Various forms

Constrained optimization (vs unconstrained)

- When there are constraints on the variables.

Discrete optimization (vs continuous)

- When variables only make sense to be discrete values.

Stochastic optimization (vs deterministic)

- When underlying model cannot be fully specified at the time of formulation.

Various forms

- Constrained optimization (vs unconstrained)
  - When there are constraints on the variables.
- Discrete optimization (vs continuous)
  - When variables only make sense to be discrete values.
- Stochastic optimization (vs deterministic)
  - When underlying model cannot be fully specified at the time of formulation.
- Convex optimization (vs nonconvex)
  - When objective and constraints are convex.

Optimization algorithms

How they operate?

- iterative: begin with an initial guess of the variable x and generate a sequence of improved estimates until they terminate, hopefully at a solution
- various strategies for moving from one iterate to the next
- can use information gathered at previous iterations
- make use of the first or second derivatives of the objective function

Optimization algorithms

Properties of good optimization algorithms:

- Robustness: They should perform well on a wide variety of problems in their class, for all reasonable values of the starting point.
- Efficiency: They should not require excessive computer time or storage.
- Accuracy: They should be able to identify a solution with precision, without being overly sensitive to errors in the data or to the arithmetic rounding errors that occur when the algorithm is implemented on a computer.

Any questions?

Credits

Numerical Optimization, Jorge Nocedal and Stephen J. Wright.