Lecture 1 Optimization

Note by Samion Suwito on 1/21/25

Definition

Optimisation is best solution to problem w/ constraints

$$p^* = \min_x f_0(x)$$
$$\text{subject to: } f_i(x) \leq 0, i = 1 \dots m$$

$x \in \mathbb{R}^n$ is the decision variable to optimize
$f_0 : \mathbb{R}^n \to \mathbb{R}$ is the objective function or cost
$f_i : \mathbb{R}^n \to \mathbb{R}, i = 1 \dots m$ is the constraints
x^* is value that minimimses f_0
p^* is optimal value = $f_0(x^*)$

We don't solve problems analytically, instead we solve problems with computers.
The *art* is to take a problem and translate it into an optimisaiotn problem

Least Squares (LS)

$\min_{x\in\mathbb{R}^n} ||Ax - b||^2$ where $A \in \mathbb{R}^{m\times n}, \vec{b} \in \mathbb{R}^m$
$x^* = (A^\top A)^{-1} A^\top b$ given A has full column rank
This is essentially

$$\sum_i (A_i^\top x + b_i)^2$$

We can transform the problem to:

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^m \lambda_i (A_i^\top x + b_i)^2 + \lambda_0 \|x - c\|^2$$
$$\downarrow$$
$$\left\| \begin{pmatrix} \lambda^{1/2} A \\ \lambda^{1/2} I \end{pmatrix} x - \begin{pmatrix} \lambda^{1/2} b \\ \lambda^{1/2} c \end{pmatrix} \right\|^2$$

In this we converted $A = \text{diag}(\lambda_1 \dots \lambda_n)$
This problem is called Regularised LS

Linear Program

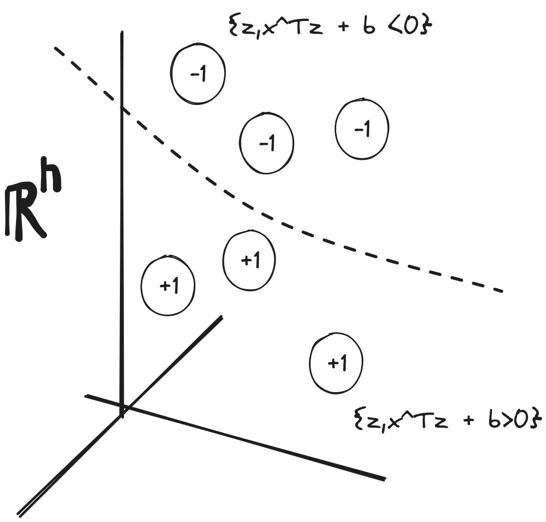
Standard Form

$$\min_{x \in \mathbb{R}^n} c^\top x$$
$$\text{s.t. } Ax \le b \;\} \; a_i^\top x \le b$$

Where A, b, c are constants that can be considered as costs that affect each component of x
Unlike LS, there are no analytical solution instead we put in the computer
$\text{linprog}(A, b, c) = x^* (\text{or } p^*)$

Example: Support Vector Machine (Classifier)

Given Data Points $z^{(i)} \in \mathbb{R}^n, i = 1 \dots m$ and labels $y_i \in \{-1, 1\}$
Goal: Find a good linear classification (hyperplane)



Where one half of the hyper plane consists of $\{z, x^\top z + b > 0\}$ and the other $\{z, x^\top z + b < 0\}$.
Find x and b

$$\min_{x \in \mathbb{R}^n, b \in \mathbb{R}} \sum_{i=1}^{m} \max(0, 1 - y_i(x^\top z^{(i)} + b))$$

Where the max function is a non linear hinge loss and we maximise with 0 because we only care whether the hyperplane classifies correctly not how far the points are from the line.

Formulation as LP:

$$\min_{x,b,r} \sum_{i=1}^{m} r_i$$

such that:

$$r_i \leq 0, \ \forall i = 1 \ldots m$$
$$r_i \geq 1 - y_i(x^\top z^{(i)} + b)$$

which is equivalent to $0 \geq 1 - y_i(x^\top z^{(i)} + b) - r_i$

Terminology

List of tricks for problem reduction
$\geq \to \leq$ by multiplying -1
Equality Constraints:

$$f_i(x) = 0 \to \begin{cases} f_i(x) \leq 0 \\ -f_i(x) \geq 0 \end{cases}$$

Set constraints

$$\min_{x \in \mathcal{X}} f_0(x) \to \min_x f_0(x) \text{ s.t. } x \in \mathcal{X}$$

Feasible - point at which constraints are met
Feasible set - set $\mathcal{X}$ where all constraints are met
Problem determined if $\mathcal{X} = \emptyset$ where $\{x, f_i(x) = 0 \text{ for } i = 1 \ldots m\}$
$\{f_i(x) \leq -1, f_i(x)\}$ is infeasible

Optimization problem is feasible if $\mathcal{X}$ is not empty
Convention: Set p^* to ∞ if infeasible
Say p^* is attained if problem feasible and exists optimal value for optimal variable x^* s.t. $f_0(x^*) = p^*$

In the following example:

$$\min_{x \geq 0} \frac{1}{x+1}$$

$p^* = 0$ but the value is not necessarily obtained
Set of optimal points = Set of optimal x^* that achieve p^*

"*tractable*" - Can be solved w/ computer in (usually) polynomial time
"*intractable*" - Cannot be solved w/ computer in reasonable time