

lecture03_Convex-Functions-and-Optimization

September 21, 2016

0.1 EECS 445 - Introduction to Machine Learning

0.2 Lecture 3: Convex Optimization and Review of Probability

0.2.1 Date: September 14, 2016

0.2.2 Instructor: Jacob Abernethy and Jia Deng

```
In [2]: from IPython.core.display import HTML, Image
        from IPython.display import YouTubeVideo
        from sympy import init_printing, Matrix, symbols, Rational
        import sympy as sym
        from warnings import filterwarnings
        init_printing(use_latex = 'mathjax')
        filterwarnings('ignore')

        %pylab inline

        import numpy as np
```

Populating the interactive namespace from numpy and matplotlib

Some important notes

- HW1 is out! Due September 26th
- Homework will be submitted via *Gradescope*. Please see Piazza for precise instructions. Do it soon, not at the last minute!!

0.3 Outline for this Lecture

- Convexity
 - Convex Set
 - Convex Function
- Introduction to Optimization
- Introduction to Lagrange Duality

In this lecture, we will first introduce convex set, convex function and optimization problem. One approach to solving an optimization problem is to solve its dual problem. We will briefly cover some basics of duality in this lecture. More about optimization and duality will come when we study support vector machine (SVM).

0.4 Convexity

0.4.1 Convex Sets

- $C \subseteq \mathbb{R}^n$ is **convex** if

$$tx + (1 - t)y \in C$$

for any $x, y \in C$ and $0 \leq t \leq 1$

- that is, a set is convex if the line connecting **any** two points in the set is entirely inside the set

(Left: Convex Set; Right: Non-convex Set)

0.4.2 Convex Functions

- We say that a function f is *convex* if, for any distinct pair of points x_1, x_2 we have

$$f(tx_1 + (1 - t)x_2) \leq tf(x_1) + (1 - t)f(x_2) \quad \forall t \in [0, 1]$$

- f is *strictly convex* if strict inequality holds when $t \in (0, 1)$
- A function f is said to be *concave* if $-f$ is convex

0.4.3 Fun Facts About Convex Functions

- If f is differentiable, then f is convex iff f “lies above its linear approximation”, i.e.:

$$f(x + y) \geq f(x) + \nabla_x f(x) \cdot y \quad \forall x, y$$

[Boyd and Vandenberghe] * If f is twice-differentiable, then f is convex iff its hessian is always positive semi-definite!

0.5 Introduction to Optimization

0.5.1 The Most General Optimization Problem

Assume f is some function, and $C \subset \mathbb{R}^n$ is some set. The following is an *optimization problem*:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array}$$

* How hard is it to find a solution that is (near-) optimal? This is one of the fundamental problems in Computer Science and Operations Research. * A huge portion of ML relies on this task

0.5.2 A Rough Optimization Hierarchy

$$\text{minimize } f(x) \quad \text{subject to } x \in C$$

* **[Really Easy]** $C = \mathbb{R}^n$ (i.e. problem is *unconstrained*), f is convex, f is differentiable, strictly convex, and “slowly-changing” gradients * **[Easyish]** $C = \mathbb{R}^n$, f is convex * **[Medium]** C is a convex set, f is convex * **[Hard]** C is a convex set, f is non-convex * **[REALLY Hard]** C is an arbitrary set, f is non-convex

0.5.3 Optimization Without Constraints

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \mathbb{R}^n \end{array}$$

* This problem tends to be easier than constrained optimization * If f is convex, we just need to find an x such that $\nabla f(x) = \vec{0}$ (necessary and sufficient) * Techniques like *gradient descent* or *Newton's method* work in this setting. (More on this later)

0.5.4 Optimization With Constraints

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_j(x) = 0, \quad j = 1, \dots, n \end{array}$$

* Here $C = \{x : g_i(x) \leq 0, h_j(x) = 0, i = 1, \dots, m, j = 1, \dots, n\}$ * C is convex as long as all $g_i(x)$ convex and all $h_j(x)$ affine (linear function + translation) * The solution of this optimization may occur in the *interior* of C , in which case the optimal x will have $\nabla f(x) = 0$ * But what if the solution occurs on the *boundary* of C ?

0.6 Introduction to Lagrange Duality

- In some cases original (**primal**) optimization problem can be hard to solve, solving a proxy problem sometimes can be easier
- The proxy problem could be **dual** problem which is transformed from primal problem
- Here is how to transform from primal to dual. For primal problem

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_j(x) = 0, \quad j = 1, \dots, n \end{array}$$

Its Lagrangian is

$$L(x, \boldsymbol{\lambda}, \boldsymbol{\nu}) := f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^n \nu_j h_j(x)$$

of which $\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\nu} \in \mathbb{R}^n$ are **dual variables**

- The **Lagrangian dual function** is

$$L_D(\boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq \inf_x L(x, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_x \left[f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^n \nu_j h_j(x) \right]$$

The minimization is usually done by finding the stationary point of $L(x, \boldsymbol{\lambda}, \boldsymbol{\nu})$ with respect to x

0.6.1 The Lagrange Dual Problem

- Then the **dual problem** is

$$\begin{aligned} &\text{maximize} && L_D(\lambda, \nu) \\ &\text{subject to} && \lambda_i, \nu_j \geq 0 \quad i = 1, \dots, m, \quad j = 1, \dots, n \end{aligned}$$

Instead of solving primal problem with respect to x , we now need to solve dual problem with respect to λ and ν

- $L_D(\lambda, \nu)$ is concave even if primal problem is not convex
- Let the p^* and d^* denote the optimal values of primal problem and dual problem, we always have *weak duality*: $p^* \geq d^*$
- Under nice conditions, we get *strong duality*: $p^* = d^*$
- Many details are omitted here and they will come when we study **support vector machine (SVM)**

0.7 Recommended reading:

- Free online!
- Chapter 5 covers duality