Lecture 06 - Convex Optimization and Regression

Instructor: Prof. Ganesh Ramakrishnan

Agenda

- Definition of Convex Sets and Functions
- Example of Convex Set
- Example of Convex Function
- Theorem related to Convex Functions
- Overfitting
- Convex Optimization Problems
- Next Lecture

Definition of Convex Sets and Convex Functions

Definition of convex sets and convex functions (Cite : Definition 32 and 35)[1]

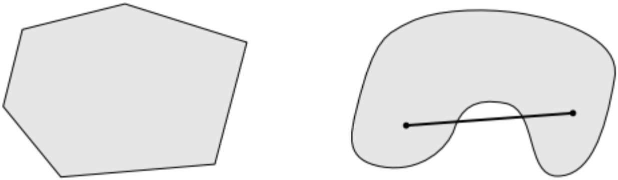

Figure: Examples of a convex set (a) and a non-convex set (b) Cite: http://cs229.stanford.edu/section/cs229-cvxopt.pdf

A set C is convex if, for any x,y $\in$ C and $\theta \in \Re$ and $0 \leq \theta \leq 1$,

$$\theta x + (1-\theta)x \in C \quad (1)$$

Example of a Convex Set

To prove : *Verify that a hyperplane is a convex set.*

Proof

- A Hyperplane $\mathcal{H}$ is defined as $\{\mathbf{x}|\mathbf{a}^T\mathbf{x} = b, \mathbf{a} \neq \mathbf{0}\}$
- Let $\mathbf{x}$ and $\mathbf{y}$ be vectors that belong to the hyperplane
- Since they belong to the hyperplane, $\mathbf{a}^T\mathbf{x} = b \textit{ and } \mathbf{a}^T\mathbf{y} = b$
- In order to prove the convexity of the set we must show that :

$$\theta\mathbf{x} + (1-\theta)\mathbf{y} \in \mathcal{H}, \ \textit{where } \theta \in [0, 1] \tag{2}$$

- In particular, it will belong to the hyperplane if it's true that :

$$\begin{aligned}
\mathbf{a}^T(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= b & (3)\\
\implies \mathbf{a}^T\theta\mathbf{x} + \mathbf{a}^T(1-\theta)\mathbf{y} &= b & (4)\\
\implies \theta\mathbf{a}^T\mathbf{x} + (1-\theta)\mathbf{a}^T\mathbf{y} &= b & (5)
\end{aligned}$$

- And, we also have $\mathbf{a}^T\mathbf{x} = b \textit{ and } \mathbf{a}^T\mathbf{y} = b$. Hence $\theta b + (1-\theta)b = b$. [Hence Proved] So a hyperplane is a convex set.

Definition of Convex Sets and Convex Functions

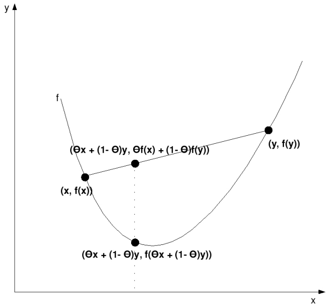

Figure: A sample convex function

$$\therefore f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}) \tag{6}$$

Example of a Convex Function

Q. *Is* $||\phi\mathbf{w} - \mathbf{y}||^2$ *convex?*

A.

- To check this, we have `(Cite : Theorem 75)`[1]. Is this practical?
- Instead, we would use `(Cite : Theorem 79)`[2] to check for the convexity of our function
- So the condition that has our focus is -

$$\nabla^2 f(\mathbf{w}^*) \text{ is positive semi} - \text{definite}, \text{ if } \forall \mathbf{x} \neq 0, \ \mathbf{x}^T \nabla^2 f(\mathbf{w}^*) \mathbf{x} \geq 0 \tag{7}$$

- We have,

$$\nabla^2 f(\mathbf{w}) = 2\phi^T\phi \tag{8}$$

- So, $\|\phi\mathbf{w} - \mathbf{y}\|^2$ is convex, since the domain for $\mathbf{w}$ is $\mathbb{R}^n$ and is convex

[1] `cs709/notes/BasicsOfConvexOptimization.pdf`

[2] `cs709/notes/BasicsOfConvexOptimization.pdf`

Strict Convexity

Q. *When is* $f(\mathbf{x})$ *(strictly) convex?*
A1. Iff $f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \ (<) \ \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y})$ for all $\theta \in [0,1]$ and for all $\mathbf{x}, \mathbf{y} \in dmn(f)$
A2. OR Iff $\nabla^2 f(\mathbf{x})$ is positive semi-definite (definite) for all $\mathbf{x} \in dmn(f)$

Strict Convexity

Q. *When is* $f(\mathbf{x})$ *(strictly) convex?*
A1. Iff $f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \ (<)\ \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y})$ for all $\theta \in [0,1]$ and for all $\mathbf{x}, \mathbf{y} \in dmn(f)$
A2. OR Iff $\nabla^2 f(\mathbf{x})$ is positive semi-definite (definite) for all $\mathbf{x} \in dmn(f)$
Q. *Is* $\|\phi\mathbf{w} - \mathbf{y}\|^2$ *strictly convex?*
A. Iff ϕ has full column rank.

To prove: *If a function is convex, any point of local minima* $\equiv$ *point of global minima*
Proof - `(Cite : Theorem 69)`[3]

[3] `cs709/notes/BasicsOfConvexOptimization.pdf`

Theorem

To prove : *If a function is strictly convex, it has a unique point of global minima*
Proof - `(Cite : Theorem 70)`[4]
Since $\|\phi\mathbf{w} - \mathbf{y}\|^2$ is strictly convex for linearly independent ϕ,

$$\nabla f(\mathbf{w}^*) = 0 \textit{ for } \mathbf{w}^* = (\phi^T\phi)^{-1}\phi^T\mathbf{y} \tag{9}$$

Thus, $\mathbf{w}^*$ is a point of global minimum. One can also find a solution to $(\phi^T\phi\mathbf{w} = \phi^T\mathbf{y})$ by Gauss elimination.

[4] `cs709/notes/BasicsOfConvexOptimization.pdf`

Redundant ϕ and Overfitting

- What do you expect in experiments if ϕ had redundancy?

Example of linearly correlated features

- Example where ϕ doesn't have a full column rank,

$$\phi = \begin{bmatrix} x_1 & x_1^2 & x_1^2 & x_1^3 \\ x_2 & x_2^2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & x_n^2 & x_n^2 & x_n^3 \end{bmatrix} \tag{10}$$

- This is the simplest form of linear correlation of features.

Redundant ϕ and Overfitting

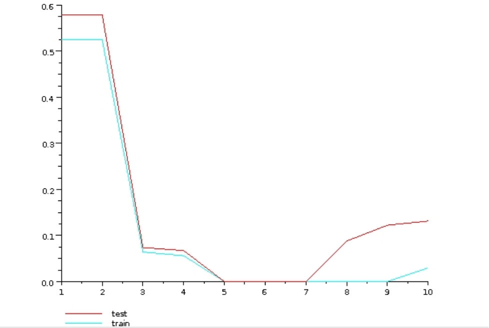

Figure: train-RMS and test-RMS values vs t(degree of polynomial) graph

- Too many bends (t=9 onwards) in curve $\equiv$ high values of some $w_i s$
- Train and test errors differ significantly