Mathematics for Machine Learning Notes

Amit Thakur

January 3, 2025

Contents

Chapter 1

Convex Analysis

A convex function is one that satisfies a specific property related to its shape.

1.1 Definition of a Convex Function

A function $f(x)$ is convex if, for any two points x_1 and x_2 in its domain, and any $\alpha \in [0, 1]$:

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) \tag{1.1}$$

This means the function value at any weighted average of two points is less than or equal to the weighted average of the function values at those two points.

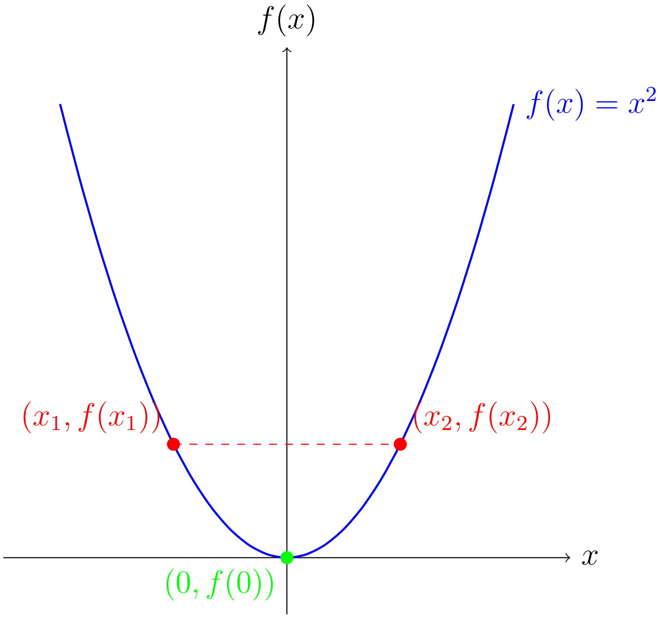

Figure 1.1: An example TikZ picture.

Chapter 2

Probability

2.1 Sample Space and Probability

2.1.1 Sets

$$
\begin{gathered}
S_1 = \{x_1, x_2, \dots, x_n\} \\
S_2 = \{x_1, x_2, \dots\} \\
S_3 = \{x \mid x \text{ satisfies } P\} \\
S_4 = \{x \mid 3 \leq x \leq 5, x \in \mathbb{R}\}
\end{gathered}
$$

- $x_2 \in S_1$.
- S_2 is an countably infinite set, as elements are enumerable.
- S_4 is

Now let's consider following sets:

$$
\begin{gathered}
S_1 = \{x_1, x_2, \dots, x_n\} \\
S_2 = \{x_1, x_2\} \\
S_3 = \{x_2, x_1\}
\end{gathered}
$$

- $S_2 \subset S_1$
- $S_2 = S_3$
- Ω: A universal set

2.1.2 Set Operations

- Complement: $S^c = \{x \in \Omega \mid x \notin S\}$
- $\Omega^c = \varnothing$ (empy set)

Chapter 3

Linear Algebra

3.1 Vectors

3.1.1 Basics of Vectors

$$\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Represent row vectors as $\mathbf{v} = \begin{bmatrix} x_1, & x_2, & \dots, & x_n \end{bmatrix}^{\mathrm{T}}$
- Zero vector: $\mathbf{0}$ has no direction.
- $\mathbf{x}^{\mathrm{TT}} = \mathbf{x}$
- Commutative: $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
- $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1, & x_2 + y_2, & \dots, & x_n + y_n \end{bmatrix}^{\mathrm{T}}$
- $\lambda \mathbf{x} = \begin{bmatrix} \lambda x_1, & \lambda x_2, & \dots, \lambda x_n \end{bmatrix}^{\mathrm{T}}$
- $(\mathbf{x} + \mathbf{y})^{\mathrm{T}} = \mathbf{x}^{\mathrm{T}} + \mathbf{y}^{\mathrm{T}}$

3.1.2 Dot Product

$$\alpha = \mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^{\mathrm{T}}\mathbf{b} = \sum_{i=1}^{n} a_i b_i \tag{3.1}$$

- Dot product value: relationship between two vectors.
- Inner Product: When two vectors are continuous functions.
- Associative property of scalar value with a dot product:

$$\lambda(\mathbf{a}^{\mathrm{T}}\mathbf{b}) = (\lambda \mathbf{a}^{\mathrm{T}})\mathbf{b} = \mathbf{a}^{\mathrm{T}}(\lambda \mathbf{b}) = (\mathbf{a}^{\mathrm{T}}\mathbf{b})\lambda$$

- Commutative property: $\mathbf{a}^{\mathrm{T}}\mathbf{b} = \mathbf{b}^{\mathrm{T}}\mathbf{a}$
- Distributive property: $\mathbf{x}^{\mathrm{T}}(\mathbf{y} + \mathbf{z}) = \mathbf{x}^{\mathrm{T}}\mathbf{y} + \mathbf{x}^{\mathrm{T}}\mathbf{z}$

Norm

$$\mathbf{a}^{\mathrm{T}}\mathbf{a} = \|\mathbf{a}\| = \sum_{i=1}^{n} a_i a_i = \sum_{i=1}^{n} a_i^2 \tag{3.2}$$

Cauchy-Schwarz Inequality

$$|\mathbf{x}^{\mathrm{T}}\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\| \tag{3.3}$$

Geometric Definition

$$\mathbf{x}^{\mathrm{T}}\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta_{xy}) \tag{3.4}$$

- $\mathbf{x} \perp \mathbf{y}$: Orthogonal if $\theta = 90^\circ = \frac{\pi}{2}$
- Collinear if $\theta = n\pi, n \in \{0, 1, 2, \ldots, N\}$. In this case, $\{\mathbf{x}, \mathbf{y}\}$ is a linearly dependent set.
- $\cos\theta$ is called the Pearson correlation coefficient.

3.1.3 Linear Weighted Combination

$$\mathbf{w} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2 + \cdots + \lambda_n\mathbf{v}_n \tag{3.5}$$

Bibliography

[1] D. Bertsekas and J.N. Tsitsiklis. *Introduction to Probability*. Athena Scientific optimization and computation series. Athena Scientific, 2008.

[2] G. Strang. *Calculus*. Number v. 1 in Open Textbook Library. Wellesley-Cambridge Press, 1991.

[3] G. Strang. *Introduction to Linear Algebra*. Wellesley, 2016.