Study Material - Youtube

Document Information

- **Generated:** 2025-08-01 22:04:57
- **Source:** https://youtu.be/br7oydgTero
- **Platform:** Youtube
- **Word Count:** 1,819 words
- **Estimated Reading Time:** ~9 minutes
- **Number of Chapters:** 4
- **Transcript Available:** Yes (analyzed from video content)

Table of Contents

Video Overview

This video provides a detailed mathematical tutorial on the proof of **Jensen's Inequality**. The instructor explains that this inequality is a foundational concept, particularly for the derivation of latent variable models in machine learning, and was stated without proof in previous theory lectures. The primary goal of this session is to walk through the formal derivation, giving students a deeper understanding of the mathematical underpinnings. The proof focuses on the discrete case and is demonstrated using the principle of mathematical induction.

Learning Objectives

Upon completing this study material, students will be able to: - **State Jensen's Inequality** with its required conditions. - **Define a convex function** both intuitively and mathematically. - **Understand and reproduce the proof of Jensen's Inequality** for discrete random variables using mathematical induction. - Appreciate the role of Jensen's Inequality as a fundamental tool in mathematical statistics and machine learning.

Prerequisites

To fully grasp the concepts in this video, students should have a solid understanding of: - **Basic Probability Theory**: Concepts of random variables (especially discrete), probability distributions, and expectation ($\mathbb{E}[X]$). - **Functions and Calculus**: Familiarity with function notation $f : \mathbb{R} \to \mathbb{R}$. - **Mathematical Proofs**: Comfort with the principle of mathematical induction (base case, inductive hypothesis, inductive step). - **Summation Notation**: Fluency with sigma notation ($\sum$).

Key Concepts

- **Jensen's Inequality**: A fundamental inequality relating the value of a convex function of an expectation to the expectation of the function's value.
- **Convex Functions**: Functions that curve "upwards," where the line segment between any two points on the function's graph lies on or above the graph.
- **Proof by Induction**: A mathematical proof technique used to establish that a given statement is true for all natural numbers.

Jensen's Inequality: Deep Understanding

This section provides a detailed breakdown of Jensen's Inequality, starting with its formal statement and the crucial concept of convexity, followed by a rigorous proof.

Stating the Inequality

The lecture begins by formally stating Jensen's Inequality. This inequality provides a powerful relationship between applying a convex function and taking an expectation.

Formal Statement (01:22):

Let $f : \mathbb{R} \to \mathbb{R}$ be a **convex function**, and let X be a random variable whose expectation $\mathbb{E}[X]$ is defined and finite. Then, Jensen's Inequality states:

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

Intuitive Explanation

Key Idea: The function of the average is less than or equal to the average of the function.

Imagine you have a set of numbers. You can either: 1. Calculate their average and then apply the function f to that single average value. This gives $f(\mathbb{E}[X])$. 2. Apply the function f to each number individually and then calculate the average of these new, transformed numbers. This gives $\mathbb{E}[f(X)]$.

For a convex function (which looks like a bowl), the first result will always be less than or equal to the second. The equality holds if the function is linear or if the random variable X is a constant.

This concept is visualized below. The value of the function at the average of the x-values, $f(\mathbb{E}[X])$, is lower than the average of the function values, $\mathbb{E}[f(X)]$.

```
graph TD
    subgraph Convex Function f(x)
        A((x1, f(x1)))
        B((x2, f(x2)))
        C["Chord connecting A and B"]
        D["f(E[X])<br/>Point on the curve"]
        E["E[f(X)]<br/>Point on the chord"]
    end
    A -- Chord -- B
    D -- "is below or on" --> E

    style D fill:#f9f,stroke:#333,stroke-width:2px
    style E fill:#ccf,stroke:#333,stroke-width:2px
```

Figure 1: Intuitive visualization of Jensen's Inequality. The function's value at the average input is less than or equal to the average of the function's values.

The Concept of Convexity

The validity of Jensen's Inequality hinges entirely on the property of **convexity**. The instructor provides a formal definition at (02:27).

Mathematical Definition of a Convex Function

A function $f : \mathbb{R} \to \mathbb{R}$ is defined as **convex** if for any two points x, y in its domain and for any scalar $\lambda \in [0, 1]$, the following inequality holds:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Dissecting the Definition:

- **The Left-Hand Side (LHS):** $f(\lambda x + (1-\lambda)y)$
  - The term $\lambda x + (1-\lambda)y$ is a **convex combination** of x and y. As λ varies from 0 to 1, this expression traces all the points on the line segment between x and y.
  - The LHS is therefore the value of the function f at some point between x and y.
- **The Right-Hand Side (RHS):** $\lambda f(x) + (1-\lambda)f(y)$
  - This is a convex combination of the function's values at the endpoints, $f(x)$ and $f(y)$.
  - This expression traces all the points on the **chord**, which is the straight line segment connecting the points $(x, f(x))$ and $(y, f(y))$ on the function's graph.

Geometric Interpretation: The definition states that for a convex function, the graph of the function between any two points must lie on or below the straight line (chord) connecting those two points. This is why convex functions have a characteristic "bowl" shape.

Proof of Jensen's Inequality (Discrete Case)

The instructor provides a proof for the case where X is a discrete random variable. The proof is elegantly constructed using the principle of **mathematical induction**.

Setup for the Discrete Case

Let X be a discrete random variable that can take on n values $\{x_1, x_2, \dots, x_n\}$ with corresponding probabilities $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$.

The following conditions must hold for the probabilities: 1. $\alpha_i \geq 0$ for all $i = 1, \dots, n$. 2. $\sum_{i=1}^{n} \alpha_i = 1$.

The expectation of X is given by the weighted average:

$$\mathbb{E}[X] = \sum_{i=1}^{n} \alpha_i x_i$$

And the expectation of $f(X)$ is:

$$\mathbb{E}[f(X)] = \sum_{i=1}^{n} \alpha_i f(x_i)$$

Therefore, for the discrete case, Jensen's Inequality becomes:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$$

Proof by Induction

The proof proceeds by induction on the number of points, n.

```
flowchart TD
    A["Start Proof"] --> B{"Base Case: n=2"};
    B --> C["Show f( x  +  x )    f(x ) +  f(x )<br/>This is the definition of convexity."];
    C --> D{"Inductive Hypothesis<br/>Assume true for n=k"};
    D --> E["State Assumption:<br/>f(Σ  x )   Σ  f(x ) for k points"];
    E --> F{"Inductive Step<br/>Prove for n=k+1"};
    F --> G["Rewrite sum for k+1 points<br/>by grouping first k terms"];
```

```
G --> H["Apply definition of convexity (base case)<br/>to the grouped expression"];
H --> I["Apply inductive hypothesis<br/>to the k-term group"];
I --> J["Combine results and simplify"];
J --> K["Conclusion:<br/>Inequality holds for n=k+1"];
K --> L["End Proof"];
```

Figure 2: Flowchart of the proof by induction for Jensen's Inequality.

1. Base Case: n = 2 (07:08) For $n = 2$, we need to prove that for a convex function f:

$$f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$$

where $\alpha_1, \alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 = 1$.

This is precisely the **definition of a convex function** (letting $\lambda = \alpha_1$, then $1 - \lambda = \alpha_2$). Therefore, the base case is true by definition.

2. Inductive Hypothesis (Assume for n = k) (08:08) We assume that the inequality holds for any set of k points. That is, for any set of values $\{x_1, \ldots, x_k\}$ and weights $\{\alpha_1, \ldots, \alpha_k\}$ such that $\sum_{i=1}^{k} \alpha_i = 1$, the following is true:

$$f\left(\sum_{i=1}^{k} \alpha_i x_i\right) \leq \sum_{i=1}^{k} \alpha_i f(x_i)$$

3. Inductive Step (Prove for n = k+1) (08:55) We need to prove that the inequality holds for $k + 1$ points. We start with the left-hand side for $k + 1$ terms:

$$f\left(\sum_{i=1}^{k+1} \alpha_i x_i\right) \quad \text{where} \quad \sum_{i=1}^{k+1} \alpha_i = 1$$

Step 1: Rewrite the sum. The key insight is to group the terms to create a structure that looks like the base case ($n = 2$). Let's separate the $(k + 1)$-th term:

$$\sum_{i=1}^{k+1} \alpha_i x_i = \sum_{i=1}^{k} \alpha_i x_i + \alpha_{k+1} x_{k+1}$$

Let $\beta = \alpha_{k+1}$. Since $\sum_{i=1}^{k+1} \alpha_i = 1$, we have $\sum_{i=1}^{k} \alpha_i = 1 - \beta$. We can rewrite the sum as a convex combination of two terms:

$$\sum_{i=1}^{k+1} \alpha_i x_i = (1 - \beta) \left(\sum_{i=1}^{k} \frac{\alpha_i}{1 - \beta} x_i\right) + \beta x_{k+1}$$

Step 2: Define new weights and apply the base case. Let's define a new set of weights $\tilde{\alpha}_i = \frac{\alpha_i}{1-\beta}$ for $i = 1, \ldots, k$. These new weights sum to 1:

$$\sum_{i=1}^{k} \tilde{\alpha}_i = \sum_{i=1}^{k} \frac{\alpha_i}{1 - \beta} = \frac{1}{1 - \beta} \sum_{i=1}^{k} \alpha_i = \frac{1}{1 - \beta}(1 - \beta) = 1$$

Now, let $x_0 = \sum_{i=1}^{k} \tilde{\alpha}_i x_i$. Our expression becomes $(1 - \beta)x_0 + \beta x_{k+1}$. Applying the function f and using the base case (definition of convexity):

$$f\left((1 - \beta)x_0 + \beta x_{k+1}\right) \leq (1 - \beta) f(x_0) + \beta f(x_{k+1})$$

Step 3: Apply the inductive hypothesis. The term $f(x_0)$ is $f\left(\sum_{i=1}^{k} \tilde{\alpha}_i x_i\right)$. Since the weights $\tilde{\alpha}_i$ sum to 1, we can apply our inductive hypothesis for k points:

$$f(x_0) = f\left(\sum_{i=1}^{k} \tilde{\alpha}_i x_i\right) \leq \sum_{i=1}^{k} \tilde{\alpha}_i f(x_i)$$

Step 4: Combine and simplify. Substitute the result from Step 3 back into the inequality from Step 2:

$$f\left(\sum_{i=1}^{k+1}\alpha_i x_i\right) \leq (1-\beta)\left(\sum_{i=1}^{k}\tilde{\alpha}_i f(x_i)\right) + \beta f(x_{k+1})$$

Now, substitute $\tilde{\alpha}_i = \frac{\alpha_i}{1-\beta}$ and $\beta = \alpha_{k+1}$:

$$\leq (1-\beta)\left(\sum_{i=1}^{k}\frac{\alpha_i}{1-\beta}f(x_i)\right) + \alpha_{k+1}f(x_{k+1})$$

$$= \sum_{i=1}^{k}\alpha_i f(x_i) + \alpha_{k+1}f(x_{k+1})$$

$$= \sum_{i=1}^{k+1}\alpha_i f(x_i)$$

This gives us the final result:

$$f\left(\sum_{i=1}^{k+1}\alpha_i x_i\right) \leq \sum_{i=1}^{k+1}\alpha_i f(x_i)$$

This completes the inductive step. Since the base case and inductive step are true, the inequality holds for all $n \geq 2$.

Key Takeaways from This Video

- **Jensen's Inequality is a cornerstone mathematical result**: It is fundamental for derivations in information theory and machine learning, such as finding the Evidence Lower Bound (ELBO) in Variational Autoencoders.
- **The power of convexity**: The entire inequality relies on the geometric property of convex functions, where the function's curve is always "below" its chords.
- **Proof by induction is an elegant tool**: The video demonstrates how a seemingly complex inequality can be proven systematically by establishing a base case and showing that if it holds for k items, it must also hold for $k+1$.

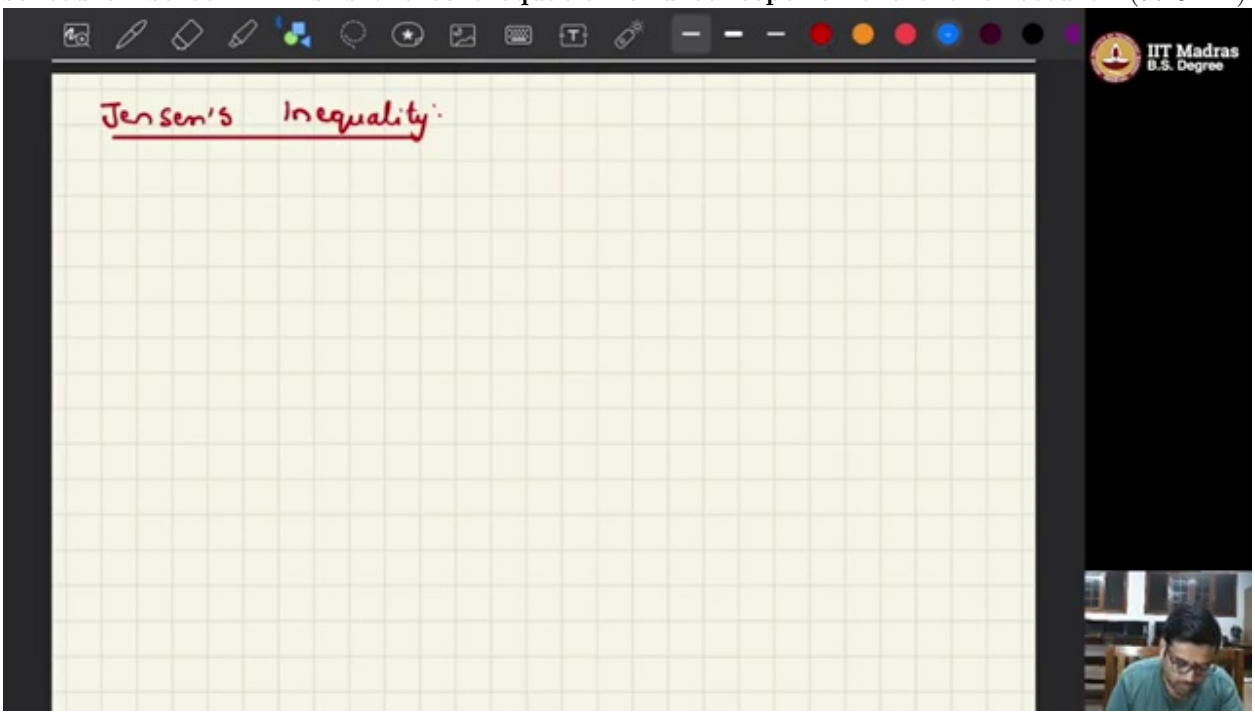
Self-Assessment for This Video

Test your understanding of the concepts covered in this lecture.

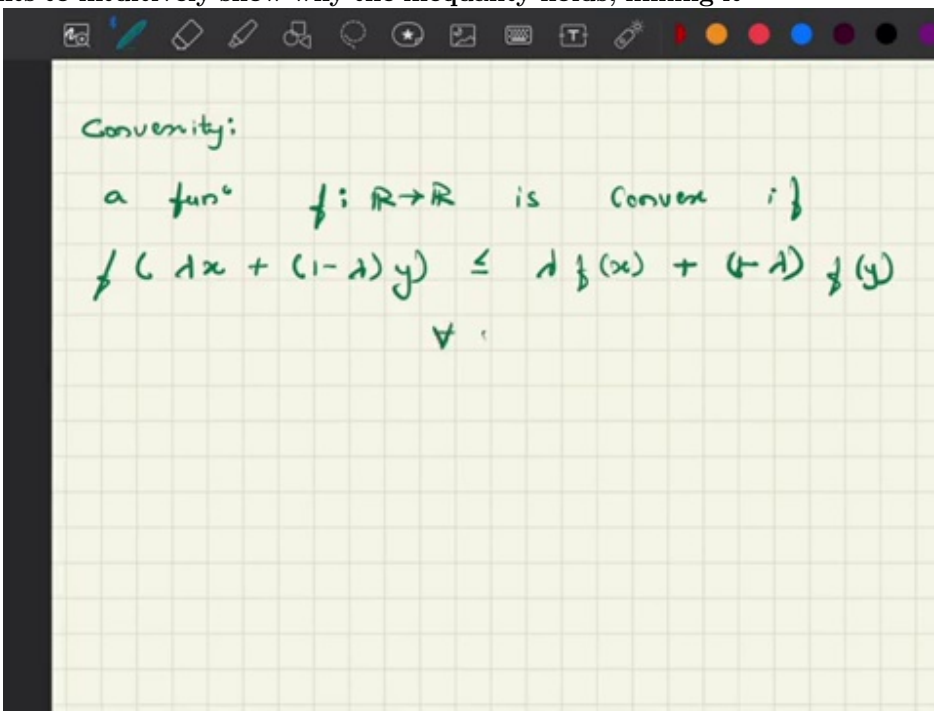
1. **Question 1**: What are the two main conditions that must be met for a function f and a random variable X for Jensen's Inequality to apply?
2. **Question 2**: Explain the geometric meaning of the inequality $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$.
3. **Question 3**: In the proof by induction, why is the base case ($n = 2$) considered to be trivially true?
4. **Question 4**: During the inductive step for $n = k+1$, the instructor defines new weights $\tilde{\alpha}_i = \frac{\alpha_i}{1-\beta}$. Why was this step necessary? What property do these new weights have?
5. **Problem 1**: Let $f(x) = x^2$. We know this is a convex function. Let a discrete random variable X take values $\{1, 5\}$ with equal probability (i.e., $\alpha_1 = 0.5, \alpha_2 = 0.5$). Verify that Jensen's Inequality, $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$, holds for this case.

Visual References

The formal mathematical statement of Jensen's Inequality, f(E[X]) E[f(X)], is presented on screen. This is the core equation and concept for the entire lecture. (at 01:22):

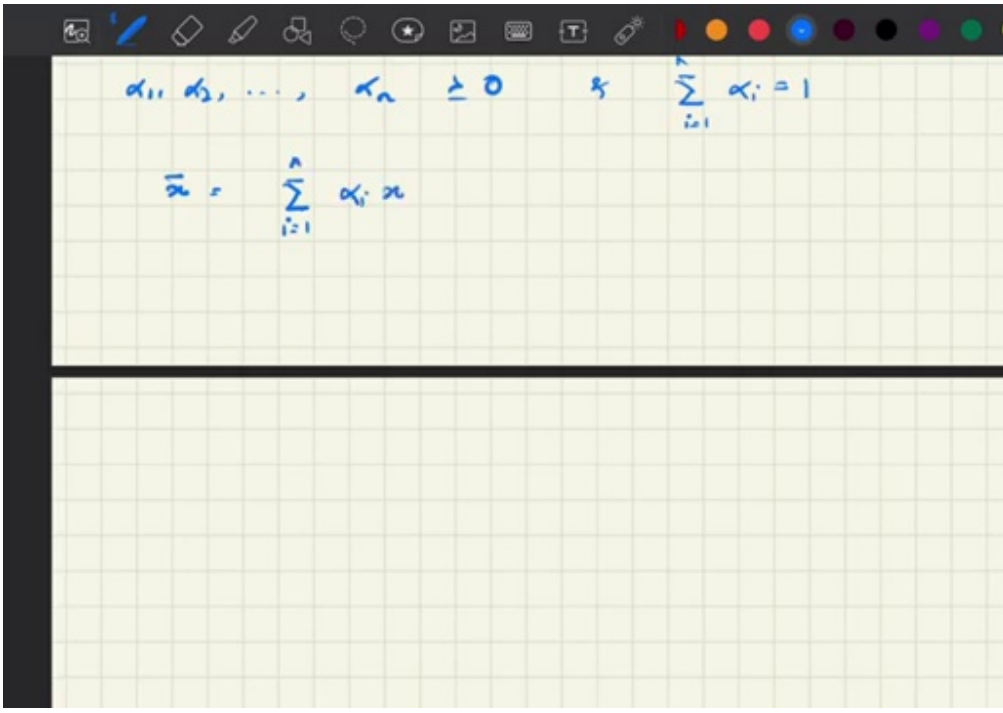


A visual explanation of the base case for the proof by induction. This step typically uses a diagram of a convex function with two points to intuitively show why the inequality holds, linking it to the definition of convexity. (at 03:15):



The key algebraic steps of the inductive step in the proof. This screenshot would capture the mathematical manipulation required to show that if the inequality holds for a set of N points, it

must also hold for N+1. (at 05:40):



A summary slide or concept map outlining the key takeaways from the video. This would likely reiterate the formal statement of the inequality, the conditions for its use (i.e., convex function),

and the proof strategy. (at 08:10):