WEEK 3: CORE MATHEMATICS FOR MACHINE LEARNING

DAY 15 (11/07/2025)

Optimization and Gradient Descent:

Once a model calculates its **loss**, the next step is to **reduce it** — to make predictions as accurate as possible. This process is called **optimization**.

Optimization is the core of machine learning: it's how the model learns from data and improves step by step.

1. What is Optimization?

- **Optimization** is about finding the best solution.
- In ML, "best" usually means **minimizing the error** (loss) between predictions and actual values.
- The model adjusts its **weights or parameters** to make the loss as small as possible.

Example:

- Predicting house prices: the model initially predicts ₹60 lakhs when the true price is ₹50 lakhs.
- Optimization changes the weights slightly so the prediction moves closer to ₹50 lakhs.
- Step by step, the model "learns" to predict more accurately.

2. Gradient Descent: Learning Step by Step

Gradient Descent is the most common optimization technique in ML.

How it works:

1. Start with random weights.
2. Calculate the loss (how wrong the predictions are).
3. Find the **gradient** — which tells the slope of the loss function with respect to each weight.
4. Adjust the weights **opposite to the gradient** to reduce loss.
5. Repeat this process many times until the loss is minimal.

3. Learning Rate: How Big Should Each Step Be?

The **learning rate** controls how much the weights change in each step:

- **Too large:** You might overshoot the minimum and never reach it.
- **Too small:** The model learns very slowly and may take too long.
- **Just right:** The model steadily approaches the minimum loss.

Example:

- If predicting student marks, a learning rate that's too high might overcorrect weights, predicting 90 instead of 80 repeatedly.
- A moderate learning rate gradually adjusts weights to get accurate predictions.

5. Intuitive Example in ML

- Predicting exam marks:
    - Initial weights = guesses for how much “hours studied” and “attendance” affect marks.
    - Loss function measures prediction errors.
    - Gradient tells **which direction to adjust weights**.
    - Step by step, the model improves predictions until error is minimized.

Reflection

Today I learned that optimization is how machines learn.

- Loss functions tell the model *how wrong it is*.
- Gradient Descent shows *how to improve step by step*.
- Learning rate controls *how fast or slow the learning happens*.

Together, they form the **learning engine of machine learning**, enabling models to gradually improve from random guesses to accurate predictions.