📄 Day 21: Variance & Gradient Descent

◆ Understanding Variance in Machine Learning

In machine learning, **variance** refers to how much the predictions of a model change when it is trained on a different dataset. It is one of the key components of the **bias-variance tradeoff**, which affects model performance.

📌 Types of Variance:

1. **Low Variance**
   - Model predictions do not change much with different training data.
   - Usually results in underfitting (high bias).
   - The model is too simple to capture the underlying patterns.
2. **High Variance**
   - Model predictions change significantly with small changes in the training data.
   - Leads to overfitting.
   - The model is too complex and captures noise along with signal.
3. **Ideal Variance**
   - The model generalizes well on unseen data.
   - Achieved by balancing bias and variance through proper regularization, model selection, and training.

◆ Gradient Descent: The Learning Engine of Neural Networks

Gradient Descent is an optimization algorithm used to minimize the loss function in machine learning models.

□ How it Works:

- It adjusts the weights (and biases) of the model iteratively.
- At each step, it moves the weights in the opposite direction of the gradient of the loss function.

□ Update Rule:

```
weight = weight - learning_rate * gradient
```

📌 Key Terms:

- **Learning Rate (α):** A small value that determines the step size during optimization.
- **Gradient:** The derivative of the loss function with respect to weights.

📊 Types of Gradient Descent:

1. **Batch Gradient Descent**
   - Uses the entire dataset for each update.
   - Stable but slow and memory-intensive.

2. **Stochastic Gradient Descent (SGD)**
    - Uses one training sample per update.
    - Faster and noisier, may converge faster.
3. **Mini-Batch Gradient Descent**
    - Uses small batches (e.g., 32 or 64 samples).
    - Balances speed and stability.
    - Commonly used in deep learning.

✅ Summary:

- Variance is a measure of model sensitivity to training data. High variance means overfitting, and low variance can lead to underfitting.
- Gradient Descent is the core optimization method used to minimize error and update weights.
- Learning rate and choice of gradient descent type significantly affect training performance and accuracy.

👓 Use in Neural Networks:

In neural networks, **gradient descent** is used with **backpropagation** to iteratively update all weights layer by layer, allowing the network to learn patterns from data and reduce prediction errors.