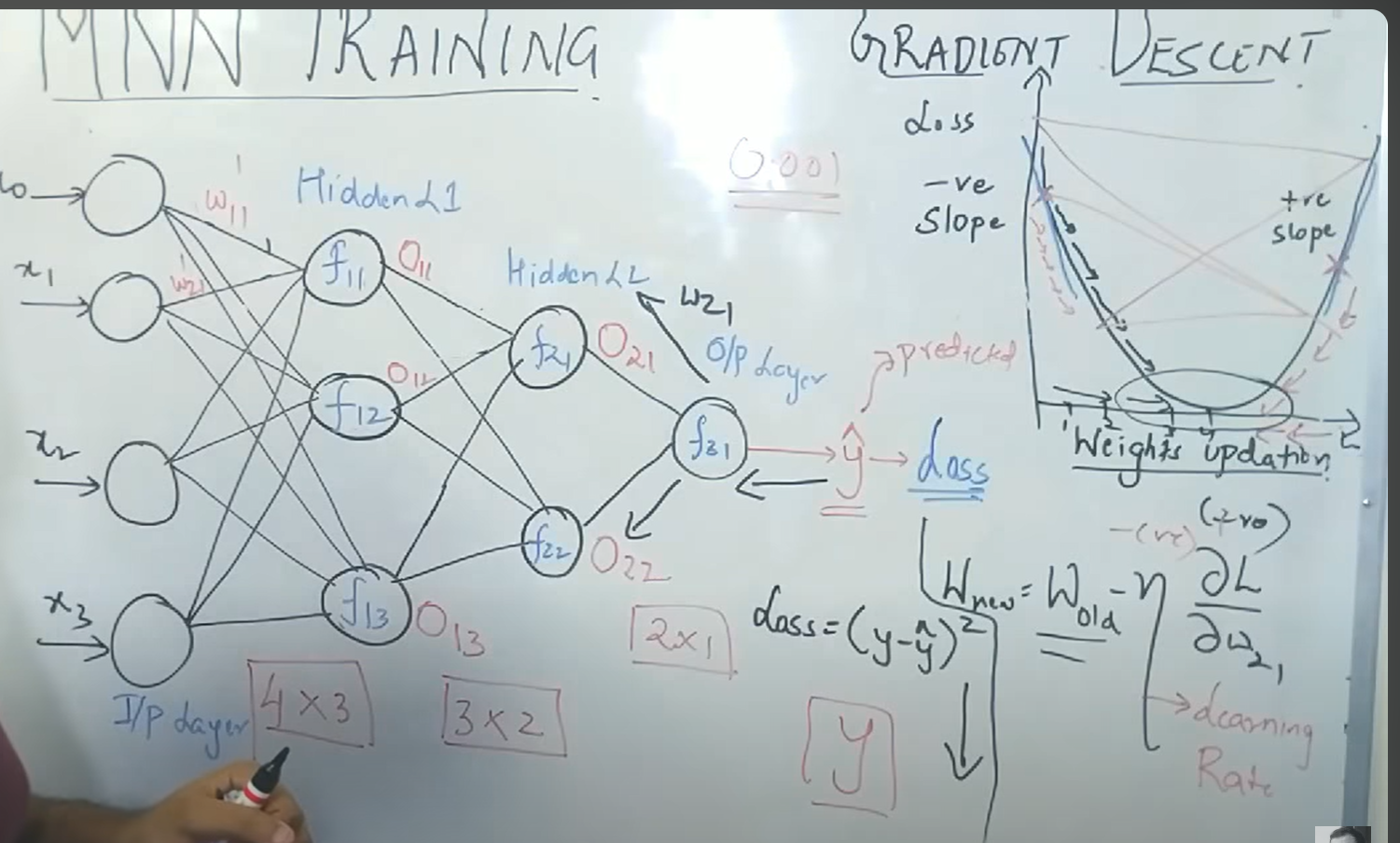
**Unit 4  
Multilayer Neural Networks and Gradient Descent**

📊 Diagram: Multilayer Neural Network and Gradient Descent



**5.1 Concept of Multilayer Neural Networks**

A multilayer neural network (MLP) consists of an input layer, one or more hidden layers, and an output layer. Each hidden layer contains neurons that transform the inputs using weights, biases, and activation functions. With multiple hidden layers, the network can learn hierarchical patterns:  
• Lower layers detect simple features (e.g., edges in images).  
• Deeper layers combine these into complex features (e.g., shapes, objects).  
This architecture makes MLPs universal function approximators capable of modeling complex non-linear relationships.

**5.2 Training of Multilayer Neural Networks**

Training involves two major phases:  
• Forward Propagation: Inputs pass through the network layer by layer to generate predicted output (ŷ).  
• Backward Propagation: The error between actual and predicted outputs is calculated. Gradients of the loss function with respect to weights are computed and used to update parameters.  
This iterative process allows the model to minimize errors and improve accuracy. Over many epochs, the model parameters converge toward optimal values.

**5.3 Concept of Gradient Descent**

Gradient descent is an optimization algorithm that minimizes the loss function by updating weights in the opposite direction of the gradient. The update rule is:  
w\_new = w\_old - η \* ∂L/∂w  
Here, η is the learning rate controlling the step size. A small η ensures stable but slow learning, while a large η speeds up training but risks overshooting minima.  
Variants of gradient descent include:  
• Batch Gradient Descent  
• Stochastic Gradient Descent (SGD)  
• Mini-batch Gradient Descent  
• Adaptive methods like Adam, RMSProp

**5.4 Loss Function and Global Minima**

The loss function measures how far predictions are from actual outputs. Common choices include Mean Squared Error and Cross-Entropy Loss.  
• Gradient descent iteratively reduces loss by adjusting weights.  
• Each step moves the weights down the slope of the loss surface.  
• Over iterations, the model converges toward a minimum point.  
• Global minima is the point where the loss is lowest possible, meaning the model is most accurate.  
In practice, models may settle in local minima or saddle points, but techniques like momentum and adaptive optimizers improve convergence.

**Summary**

• Multilayer neural networks contain multiple hidden layers that capture hierarchical representations.  
• Training involves forward propagation for predictions and backward propagation for weight updates.  
• Gradient descent is the foundation of optimization, guiding weights to reduce loss.  
• Loss functions measure prediction error; gradient descent iteratively moves the model toward global minima.  
• Together, MLPs and gradient descent underpin most modern deep learning architectures.