◾ Activation Function:

➢ $A = \alpha(W^T X + b)$

In this formula,

- A represents the activation of the neuron.
- α is the activation function applied to the weighted sum of inputs.
- W^T denotes the transpose of the weight matrix W, which represents the synaptic weights connecting the neuron to its inputs.
- X is the input vector.
- b is an optional bias term.

➢ In this way, we calculate the output of each layer.
➢ RELU activation function is good one to start with.

➢ Ani output layer ko activation function chai kasto type ko output chayinxa tesma bhar parxa.

◾ Batch size:

➢ Batching vaneko data lai small pieces maa separate garnu ho.
➢ Processors can run through them more quickly
➢ Helps in Generalization too

➢ Yo gradient descent maa badhi sunninxa / use hunxa.

➢ Here are few types of it:-

1. **Batch Gradient Descent**: It's like learning from all your data at once, which is accurate but can be slow.
2. **Stochastic Gradient Descent (SGD)**: It's like learning from one example at a time, which is fast but can be noisy.
3. **Mini-batch Gradient Descent**: It's like learning from a small group of examples at a time, which balances speed and accuracy in learning.

◾ Iterations / Epochs :

➢ Yo chahi model ko performance maa bhar parxa

➢ Early stopping is the way to determine the optimal number.

(Early stopping technique use garera hamle model lai halt garna sakinxa ,when model ko performance degrade huna thalxa inorder to prevent the overfitting.)