

Question 1: Image Classification from Scratch

In the designed architecture, each CIFAR-10 image is flattened into 3072 input features ($32 \times 32 \times 3$) and passed to a fully connected layer with 128 nodes. The net input to this layer is computed as a weighted sum and passed through the ReLU activation function. The output of the ReLU is then passed to the next fully connected layer with 3 nodes, where a Softmax activation function is applied to compute the probabilities for each of the three classes.

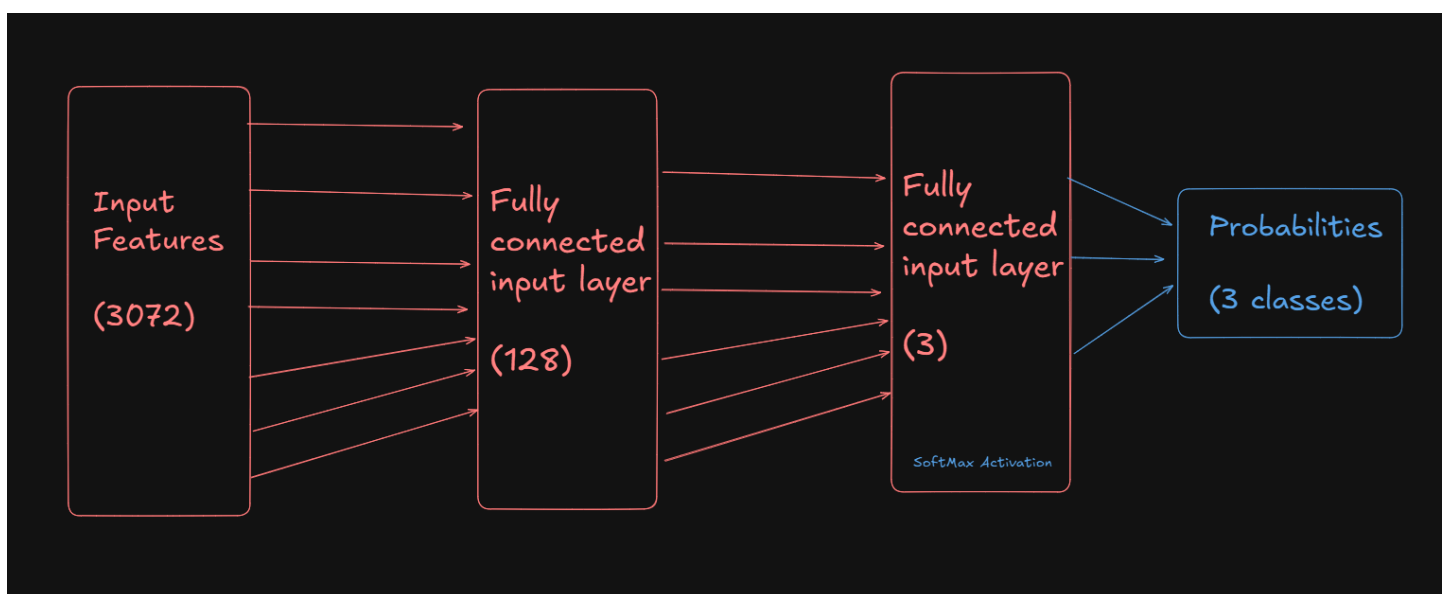
After the forward pass, the model computes the loss (error) using the cross-entropy loss function. This loss is then used to compute the gradients via backpropagation. The gradients of the loss with respect to the weights and biases are calculated and used to update the weights using gradient descent to improve the model's performance.

This process of forward pass \rightarrow loss computation \rightarrow backpropagation \rightarrow weight update is repeated for each mini-batch of the training data within an epoch. After all mini-batches in an epoch are processed, a forward pass on the test set is performed to evaluate accuracy. The entire training loop is repeated for 100 epochs, allowing the model to learn and generalize better over time.

ReLU function: $f(x) = \max(0, x)$

SoftMax function: $f(z) = \exp(z) / \sum(\exp(z_i))$; [Sum of all = 1]

To make the model more robust, different activation functions like Leaky ReLU, SiLU or even GeLU can be applied. Leaky ReLU will solve the dying ReLU problem and make sure the weight are updating more efficiently. SiLU and GeLU are better than ReLU in some cases because they are smooth and non-monotonic, allowing better gradient flow and improved performance in deep networks.



Architecture of the neural network