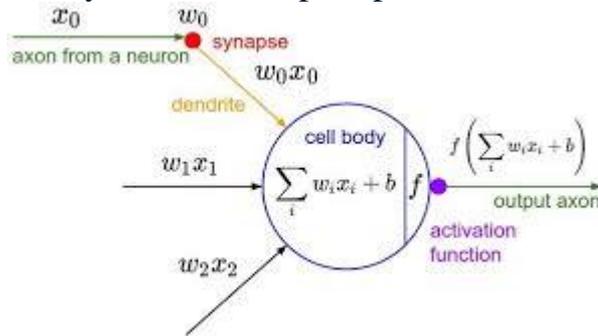




Hyper parameter–Activation Function

In deep learning, the activation function is a crucial hyperparameter that determines how a neuron's output is transformed, significantly impacting model performance and the ability to learn complex patterns.



Here's a more detailed explanation:

What are Hyperparameters?

- Hyperparameters are settings that control the learning process of a machine learning model, set before training begins, and are not learned from the data itself.
- Examples include the number of layers, nodes per layer, learning rate, and the chosen activation function.
- Properly configured hyperparameters are essential for achieving a balance between underfitting and overfitting, leading to better generalization on unseen data.

What is an Activation Function?

- An activation function introduces non-linearity into a model, enabling it to learn complex, non-linear relationships in the data.
- It transforms the output of a neuron (the weighted sum of its inputs) into a value that can be passed to the next layer.
- Common activation functions include ReLU, Sigmoid, Tanh, and Softmax.
- The choice of activation function can significantly impact the model's ability to learn and its performance.
- For example, ReLU is often used in hidden layers to address the vanishing gradient problem, while Sigmoid and Softmax are used in output layers for binary and multi-class classification, respectively.

Why is it a Hyperparameter?

19CST302&Neural Networks and Deep Learning

S.VASUKI



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

- The activation function is a setting that you, the data scientist, choose before training, not something the model learns from the data.
- Different activation functions have different characteristics and can lead to different training dynamics and performance outcomes.
- Therefore, the choice of activation function is a hyperparameter that needs to be tuned to optimize model performance for a specific task and dataset.

Examples of Activation Functions:

- **ReLU (Rectified Linear Unit):**

Outputs the input directly if it's positive, otherwise, outputs 0.

- **Sigmoid:**

Outputs a value between 0 and 1, often used in binary classification output layers.

- **Tanh:**

Outputs a value between -1 and 1, similar to sigmoid but with a wider range.

- **Softmax:**

Outputs a probability distribution over multiple classes, used in multi-class classification output layers.

