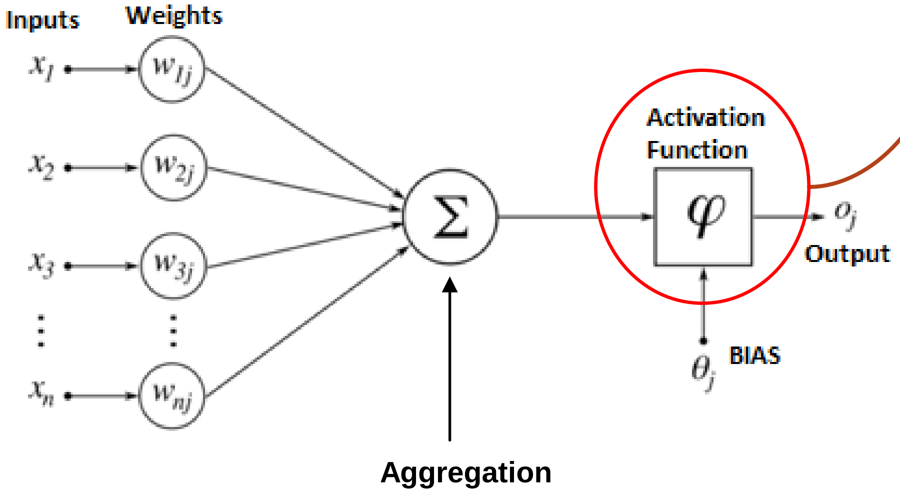
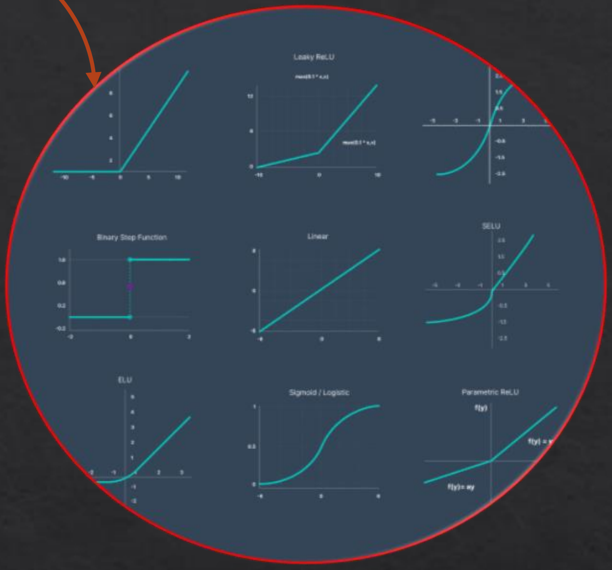
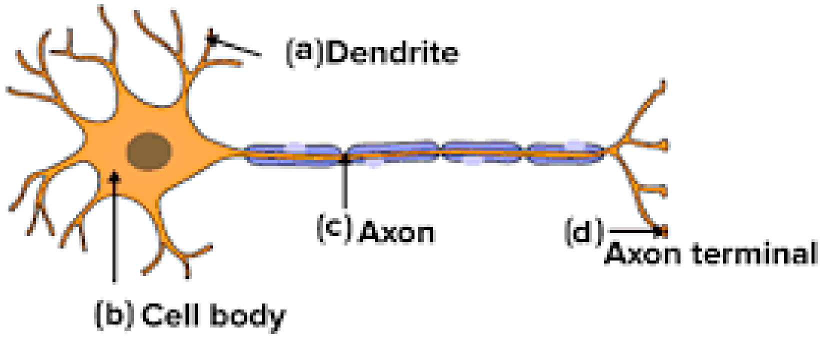
Artificial Neural Networks

Basic MLP for Classification





What are them?



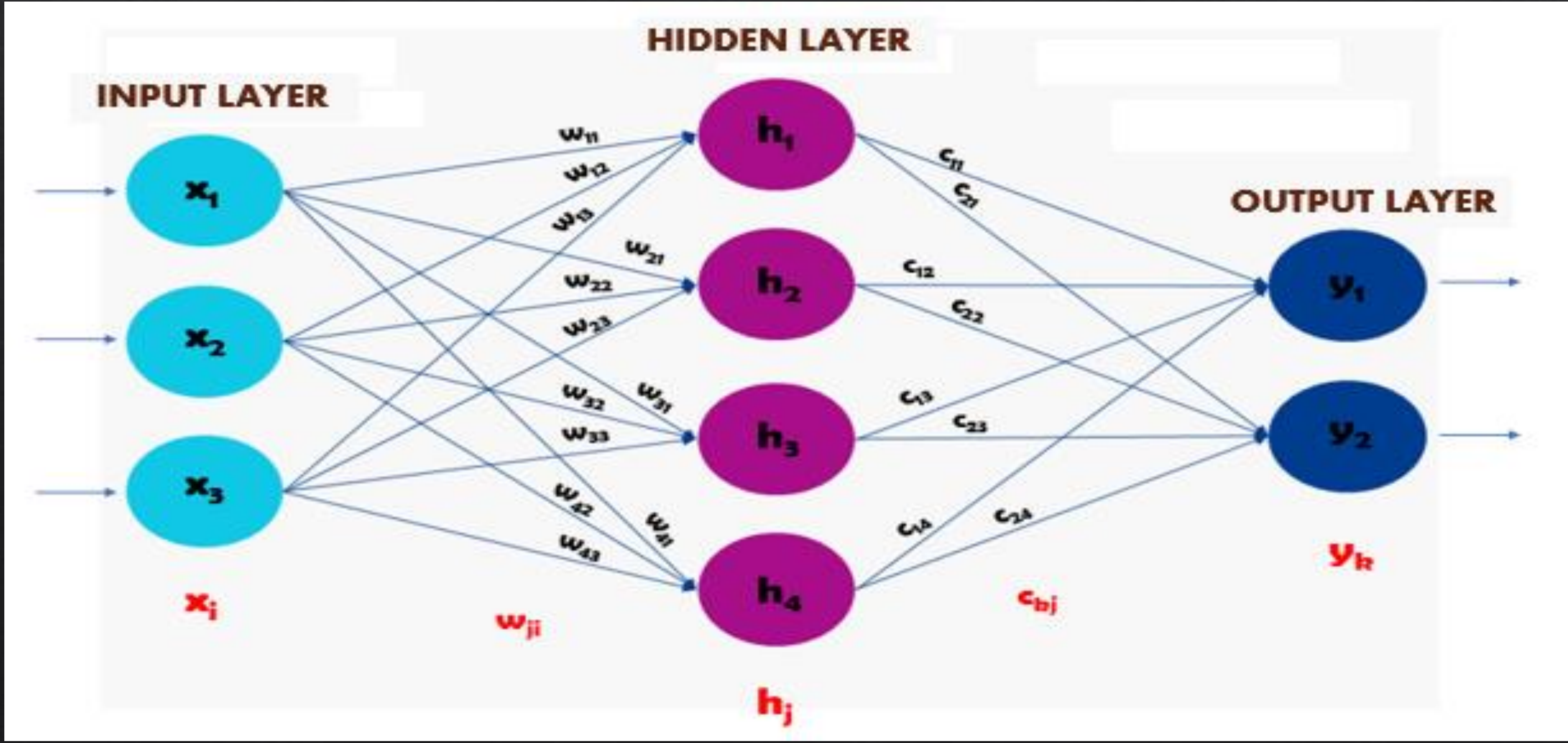
- Machine Learning models inspired by the workings of the human brain
- They are made up of layers of interconnected nodes, called neurons, which process input information to produce an output.

$$A_h = wx + b \leftrightarrow \begin{bmatrix} A_{h_1} \\ A_{h_2} \\ \vdots \\ A_{h_j} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1i} \\ w_{21} & w_{22} & \cdots & w_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ w_{j1} & w_{j2} & \cdots & w_{ji} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_i \end{bmatrix} + b$$

$$A_y = ch + b \leftrightarrow \begin{bmatrix} A_{y_1} \\ \vdots \\ A_{y_k} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1j} \\ c_{21} & c_{22} & \cdots & c_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{kj} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_j \end{bmatrix} + b$$

$$h = \zeta(A_h)$$

$$y = \zeta(A_y)$$



*Depend on the Data

*Ideally ‘0 < b < 1’

How are they trained?

Dataset Preprocessing

Split Dataset in Training and Validation

Define the ANN Topology

- **FeedForward**
- Recurrent
- Convolutional
- Autoencoder
- LSTM
- Generatives

Define Training Parameters

- Learning Rate
- Epochs / Iterations
- Layer Configurations (Activation Functions)
- Loss Function
- Optimizer

Training / Fitting

Validation

Model Prediction

Generalized Delta Rule

$$w = w - \alpha \cdot \frac{\partial ECM}{\partial w}$$

$$c = c - \alpha \cdot \frac{\partial ECM}{\partial c}$$

Mean Squared Error

$$ECM = \frac{1}{2}\sum (Y_D - Y_k)^2$$

- Absolute Error
- Hubber Loss
- Poisson Loss
- Cross Entropy
- Logarithmic Error

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\sigma'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$\sigma'(x) = \sigma(x) \cdot (1 - \sigma(x))$$

Sigmoid Activation Function

- ReLu
- Hyperbolic Tangent
- SoftMax
- Exponential
- Swish

$$\frac{\partial ECM}{\partial c_{kj}} = \frac{\partial ECM}{\partial Y_k} \cdot \frac{\partial Y_k}{\partial A_y} \cdot \frac{\partial A_y}{\partial c_{kj}} = \left[\frac{\cancel{2}\,\overset{e_k}{(Y_D - Y_k)}}{\cancel{2}} \cdot (-1)\right] \cdot [Y_k \cdot (1 - Y_k)] \cdot h_j \quad \Rightarrow \boldsymbol{c_{kj} = c_{kj} + \alpha \cdot e_k \cdot [Y_k \cdot (1 - Y_k)] \cdot h_j}$$

$$\frac{\partial ECM}{\partial w_{ji}} = \frac{\partial ECM}{\partial Y_k} \cdot \frac{\partial Y_k}{\partial A_y} \cdot \frac{\partial A_y}{\partial h_j} \cdot \frac{\partial h_j}{\partial A_h} \cdot \frac{\partial A_h}{\partial w_{ji}} = \left[\frac{\cancel{2}\,\overset{e_k}{(Y_D - Y_k)}}{\cancel{2}} \cdot (-1)\right] \cdot [Y_k \cdot (1 - Y_k)] \cdot c_{kj} \cdot [h_j \cdot (1 - h_j)] \cdot x_i$$

$$\Rightarrow \boldsymbol{w_{ji} = w_{ji} + \alpha \cdot e_k \cdot [\mathrm{Y_k} \cdot (1 - \mathrm{Y_k})] \cdot \mathrm{c_{kj}} \cdot [\mathrm{h_j} \cdot (1 - \mathrm{h_j})] \cdot x_i}$$