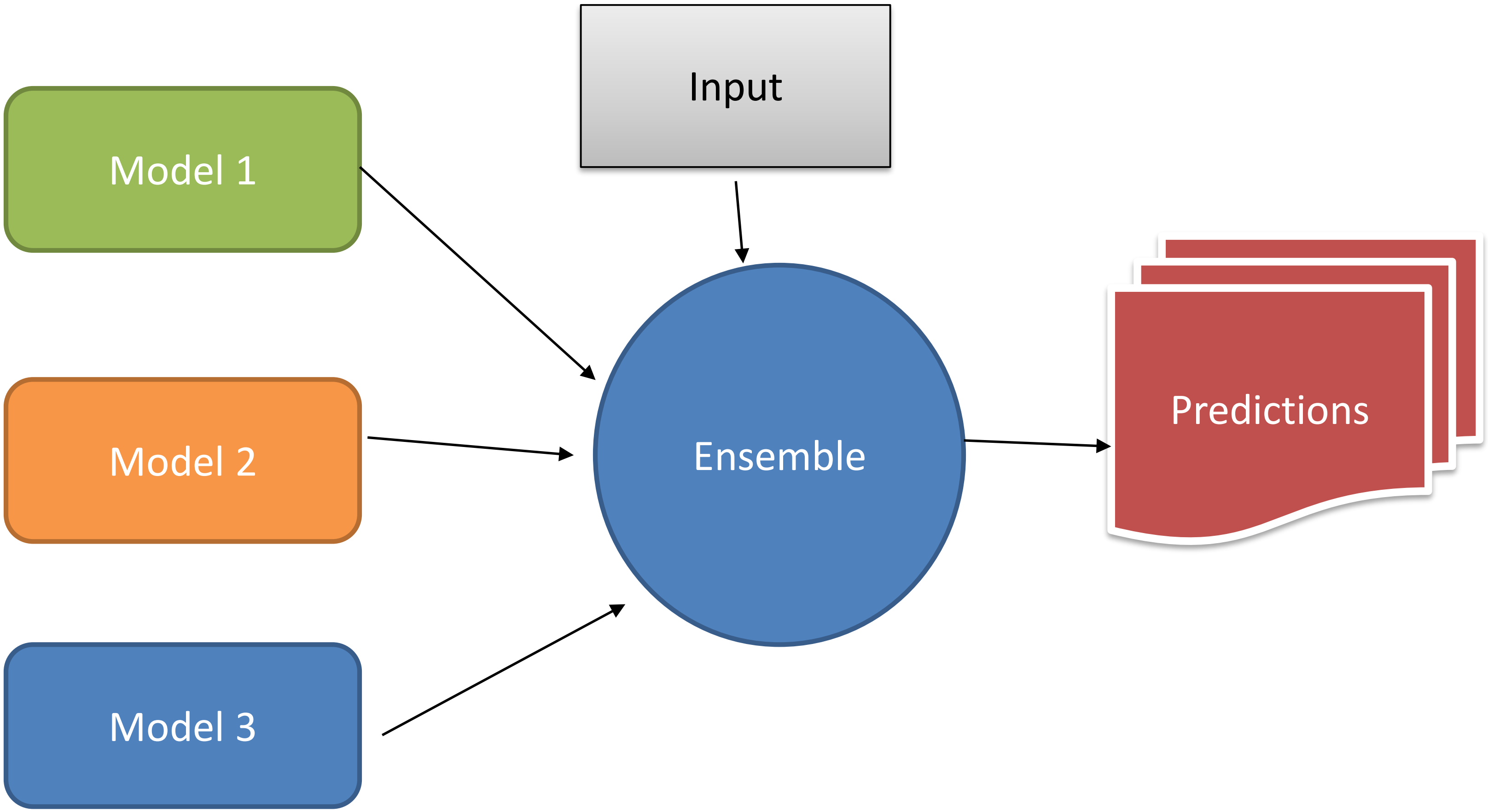
TRENDS IN MACHINE LEARNING

- Ensemble Learning: Combining Multiple Models, Bagging, Randomization, Boosting, Stacking
- Reinforcement Learning: Exploration, Exploitation, Rewards, Penalties
- Deep Learning: The Neuron, Expressing Linear Perceptron as Neurons, Feed Forward Neural Networks, Linear Neurons and their Limitations, Sigmoid, Tanh and ReLU Neurons

Ensemble

- Powerful way to improve the performance of the model by combining output of multiple classifier
- Combinations of models are known as Model Ensembles
- Increases algorithmic and model complexity


Input
Model 1
Model 2
Model 3
Ensemble
Predictions

- Lower error
- Less overfitting

Ensemble Methods

- Bagging
  - Decrease variance
- Boosting
  - Decrease bias
- Stacking
  - Improve projections

Bagging

- Bootstrap aggregating
- Creates diverse models of different random samples of the original data set.
- The samples are taken uniformly with replacement .These samples are know as bootstrap samples.

Bagging

- Combining the various outputs into single prediction (Weak learners)
  - Take a weighted Vote
  - Average /aggregate
- Increases stability and accuracy of the model by reducing the variance

Algorithm 11.1: Bagging($D, T, \mathscr{A}$) – train an ensemble of models from bootstrap samples.

Input : data set D; ensemble size T; learning algorithm $\mathscr{A}$.

Output : ensemble of models whose predictions are to be combined by voting or averaging.

for $t = 1$ to T **do**

  build a bootstrap sample D_t from D by sampling $|D|$ data points with replacement;

  run $\mathscr{A}$ on D_t to produce a model M_t;

end

return $\{M_t | 1 \leq t \leq T\}$

Combining the result

- Voting
  - Majority of the class wins
- Averaging
  - Mean of all predictions

Boosting

- Iterative technique
- Adjust the weights based on last classification.
- Reduces bias
- Convert weak learners to strong one

Boosting

- Algorithm AdaBoost

Given: $(x_1, y_1), \dots, (x_m, y_m)$ where $x_i \in X, y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$.
For $t = 1, \dots, T$:

- Train weak learner using distribution D_t.
- Get weak classifier $h_t: X \to \mathbb{R}$.
- Choose $\alpha_t \in \mathbb{R}$.
- Update:

$$D_t + 1(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

Where Z_t is a normalization factor

$$Z_t = \sum_{i=1}^{m} D_t(i) exp\,(-\alpha_t y_i h_t(x_i))$$

Output the final classifier:

$$H(x) = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$

Boosting

Choose α_t to minimize training error

$$\alpha_t = \frac{1}{2} In\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

where

$$\epsilon_t = \sum_{i=1}^{m} D_t(i)\delta(h_t(x_i) \neq y_i)$$

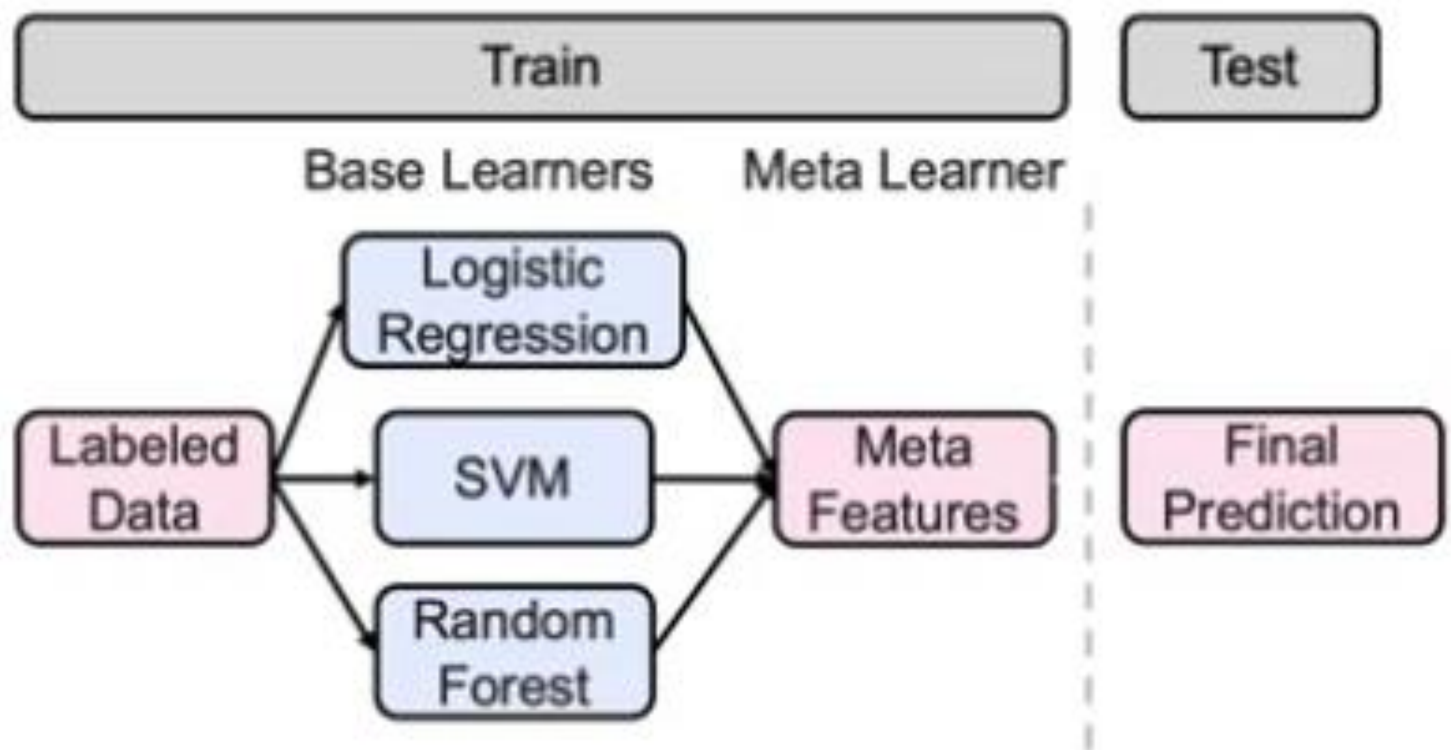
Algorithm 11.3: Boosting($D, T, \mathscr{A}$) – train an ensemble of binary classifiers from reweighted training sets.

Input : data set D; ensemble size T; learning algorithm $\mathscr{A}$.
Output : weighted ensemble of models.

$w_{1i} \leftarrow 1/|D|$ for all $x_i \in D$; // start with uniform weights
for $t = 1$ to T **do**
    run $\mathscr{A}$ on D with weights w_{ti} to produce a model M_t;
    calculate weighted error ϵ_t;
    if $\epsilon_t \geq 1/2$ **then**
        set $T \leftarrow t - 1$ and break
    end
    $\alpha_t \leftarrow \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$; // confidence for this model
    $w_{(t+1)i} \leftarrow \frac{w_{ti}}{2\epsilon_t}$ for misclassified instances $x_i \in D$; // increase weight
    $w_{(t+1)j} \leftarrow \frac{w_{tj}}{2(1-\epsilon_t)}$ for correctly classified instances $x_j \in D$; // decrease weight
end
return $M(x) = \sum_{t=1}^{T} \alpha_t M_t(x)$

Stacking

- Stacked generalization
- less widely used than bagging and boosting
- applied to models built by different learning algorithms
- Uses the concept of a meta-learner, which replaces the voting procedure


Train
Test
Base Learners
Meta Learner
Logistic Regression
Labeled Data
SVM
Meta Features
Final Prediction
Random Forest

Reinforcement Learning

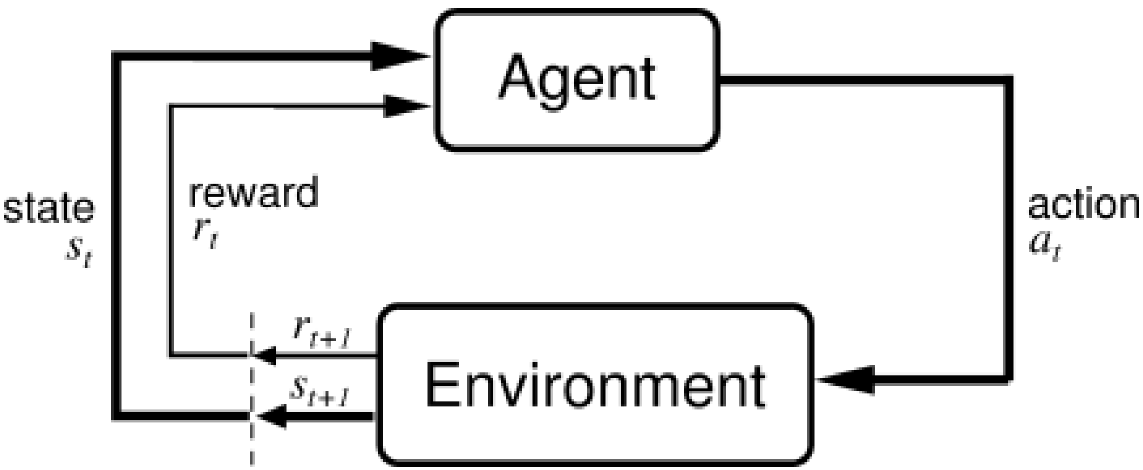
- In this approach learners or software agents learn from direct interaction with the environment.
- Agent gets a feedback about the actions as reward or punishment.
- Combines the field of dynamic programming and supervised learning

Reinforcement Learning

- Exploitation
  - Making the best use of knowledge acquired so far
- Exploration
  - Exploring new action
- Each action leads to learning through rewards or penalties

Reinforcement Learning

- what to do and how to map situations to actions to maximize the numerical reward signal.



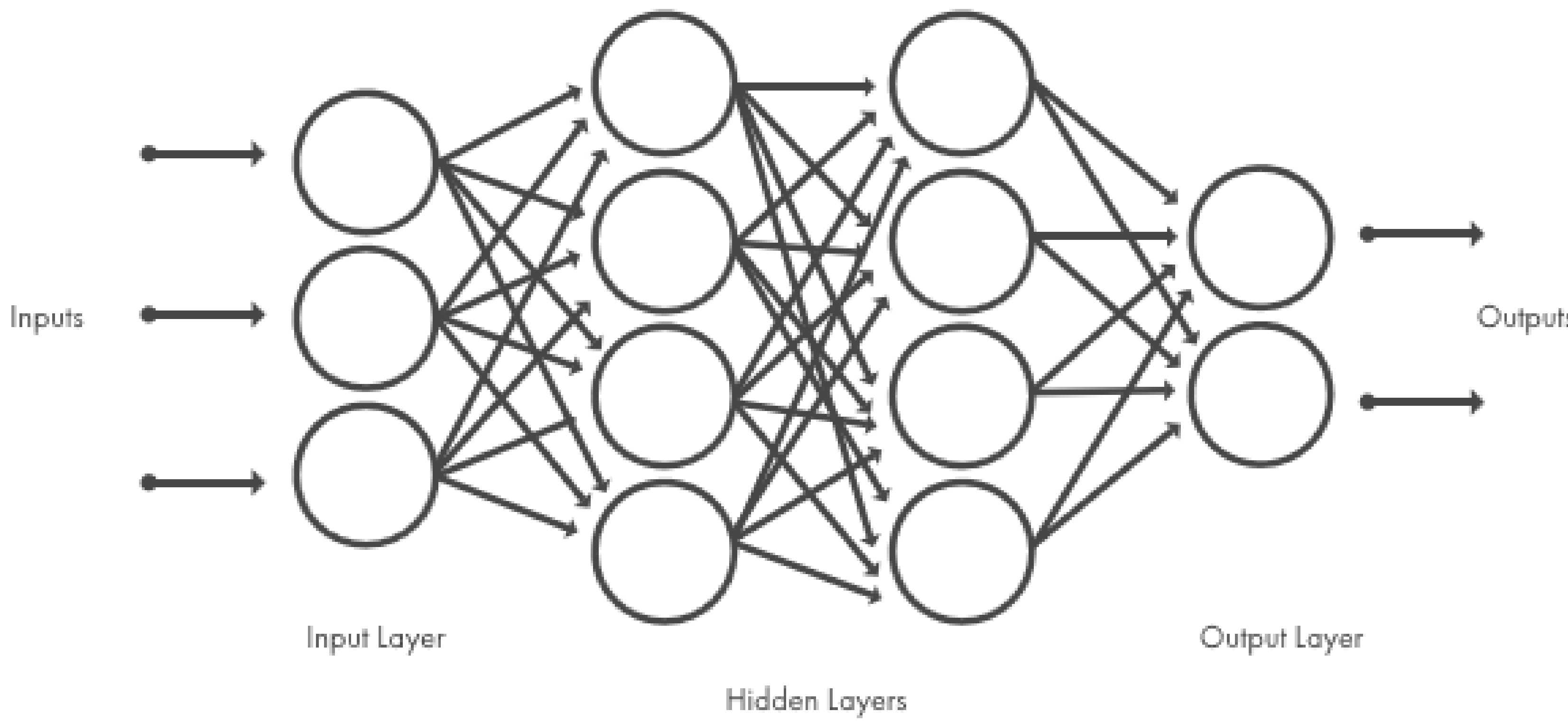
$$V^{\pi}(s) = E_{\pi}\{\, r_{t+1} + \gamma r_{t+2} + \gamma^{2}\, r_{t+3} + \ \dots /\ \ s_t = s\}$$

Deep Learning

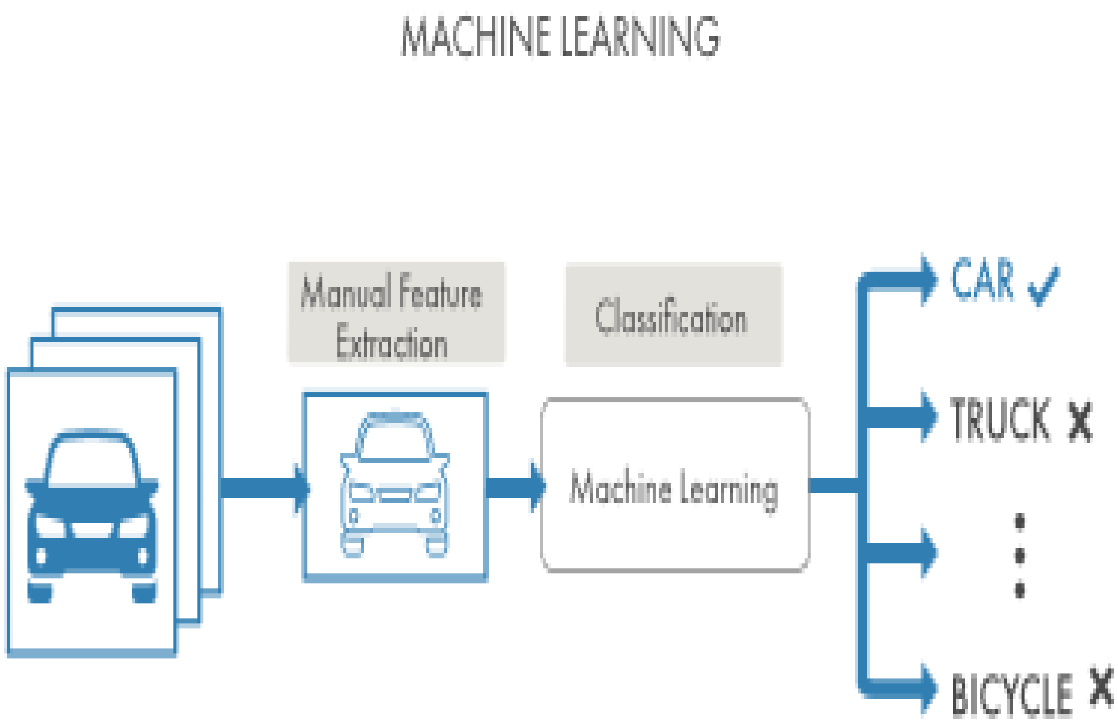
- Subfield of Machine Learning
- Deep learning is a machine learning technique that teaches computers to learn by example.
- Examples
  - Automated Driving
  - Industry Automation
  - Medical Research
  - Object recognition

- requires large amounts of **labeled data**
- requires substantial **computing power**
- Most deep learning methods use **neural network** architectures, which is why deep learning models are often referred to as **deep neural networks**.

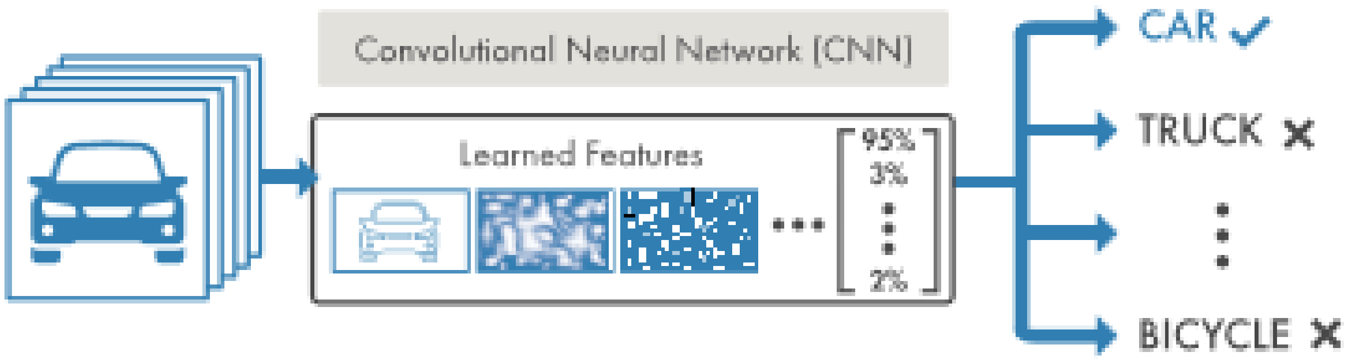
Neural Network



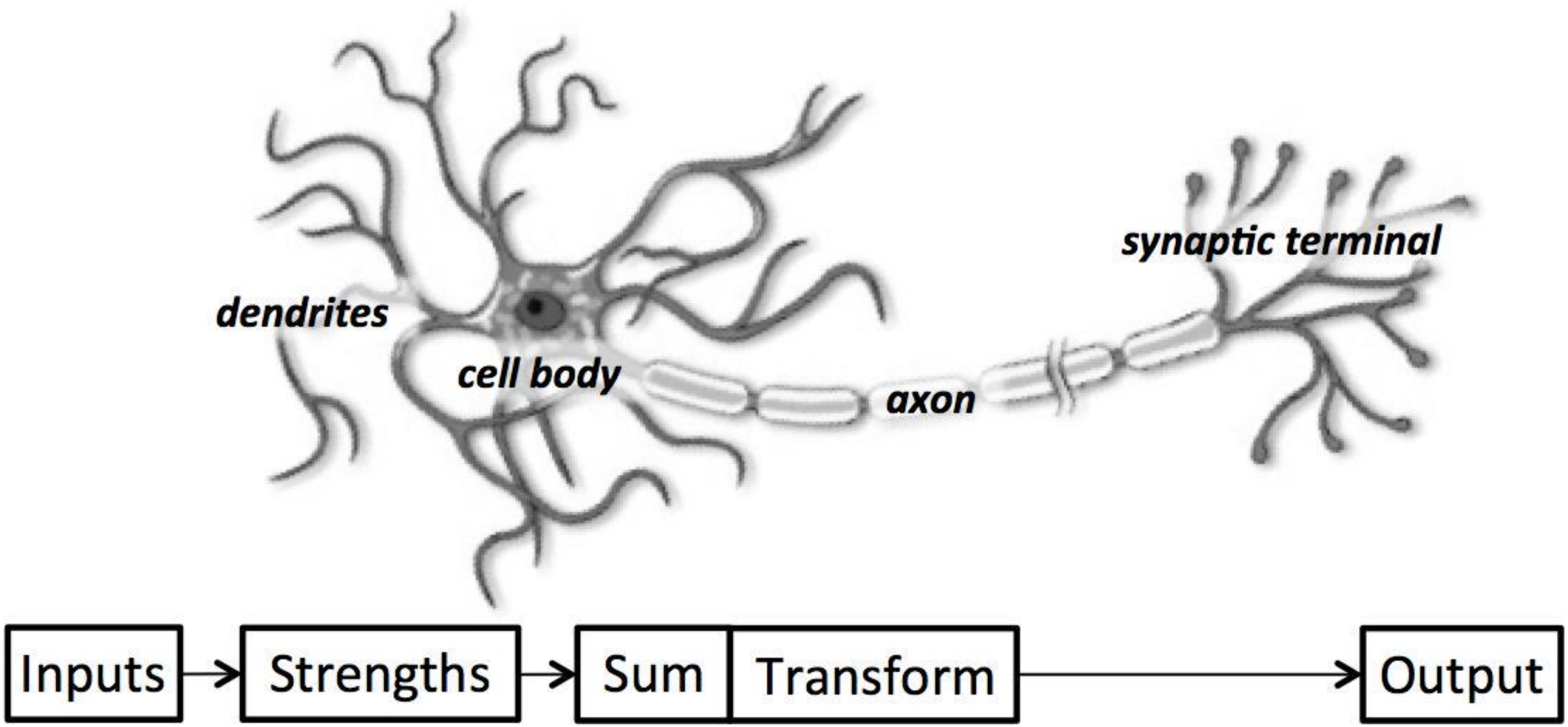
Difference Between Machine Learning and Deep Learning



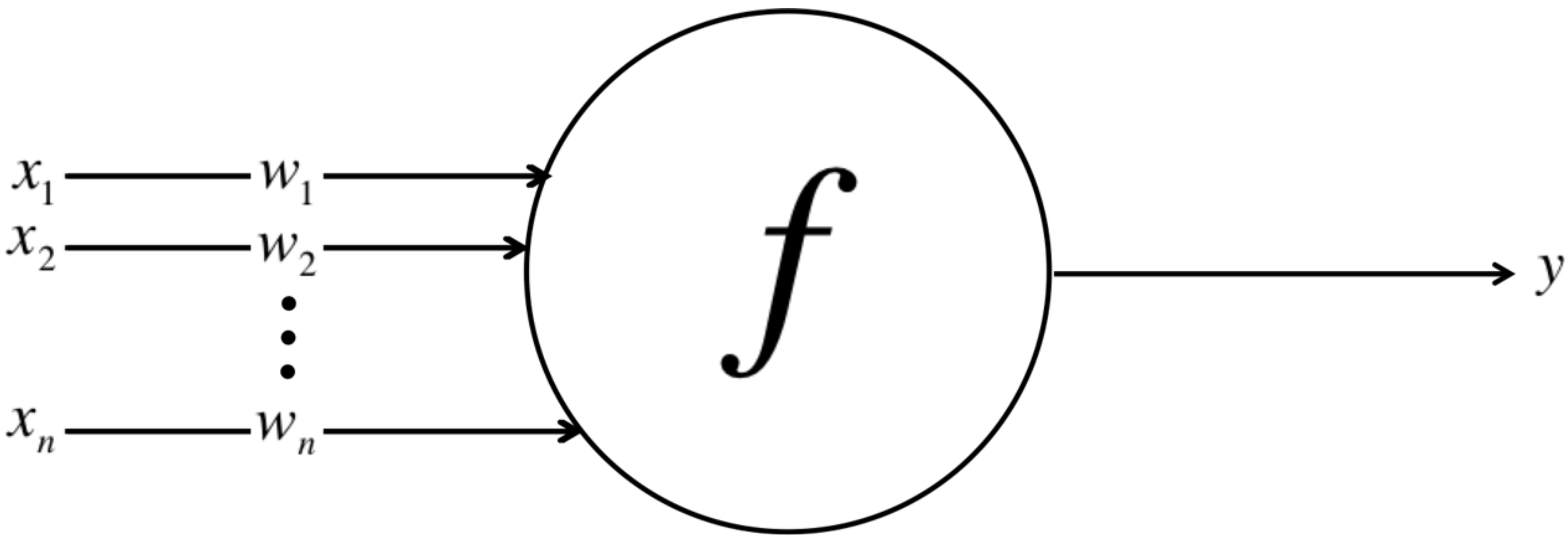
DEEP LEARNING



Neuron's functional Structure

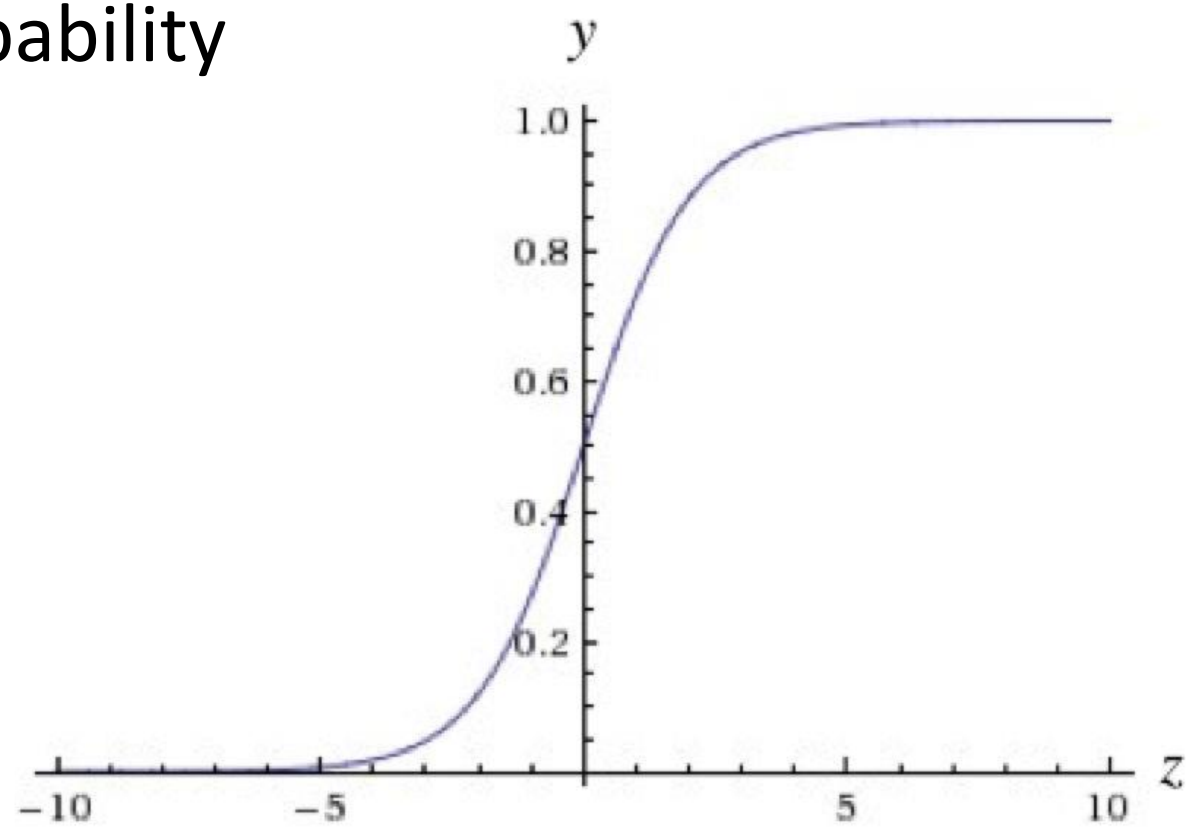


Neuron in ANN



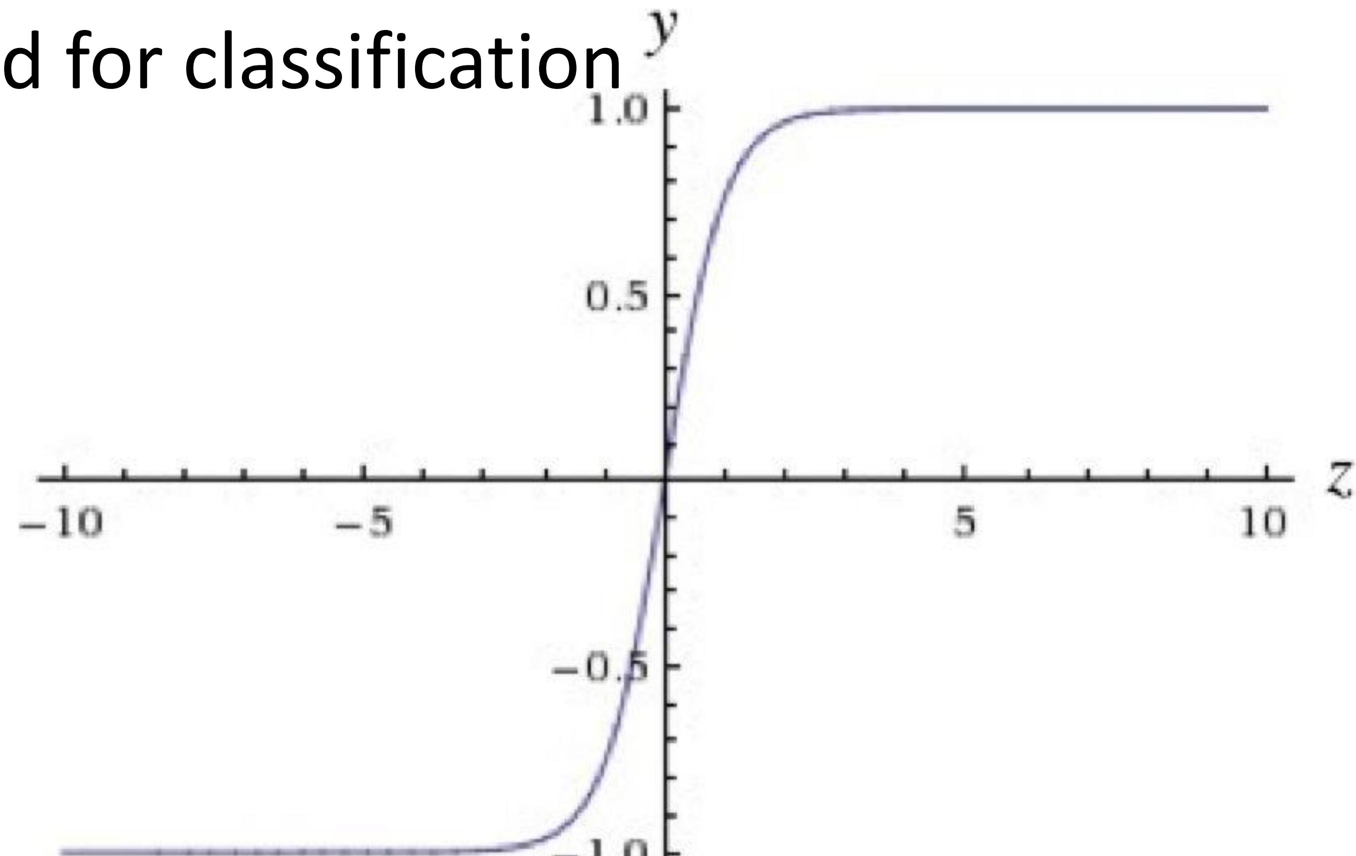
Types of Neurons

- Sigmoid $f(z) = \frac{1}{1+e^{-z}}$
- The out put is between 0 to 1
- Predict the probability



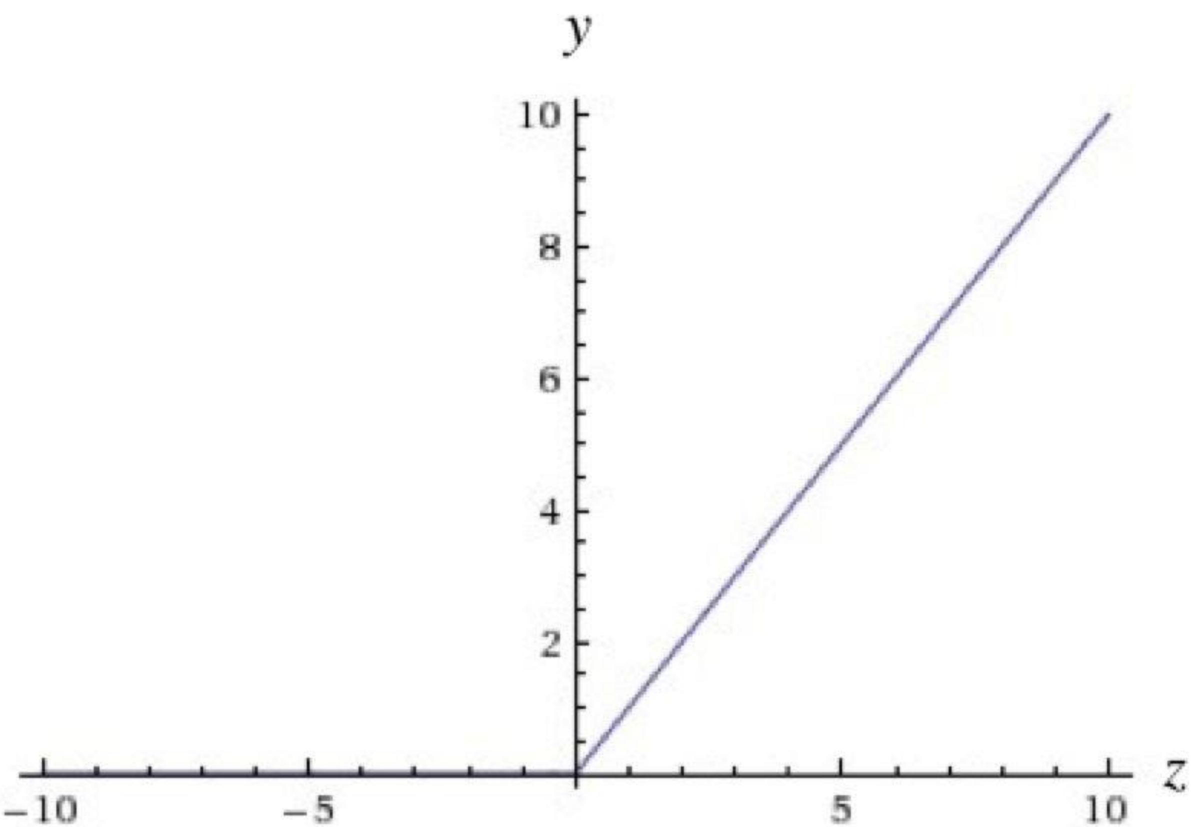
Types of Neurons

- Sigmoid $f(z)$ =tanh(z)
- output of tanh neurons range from −1 to 1
- Mainly used for classification



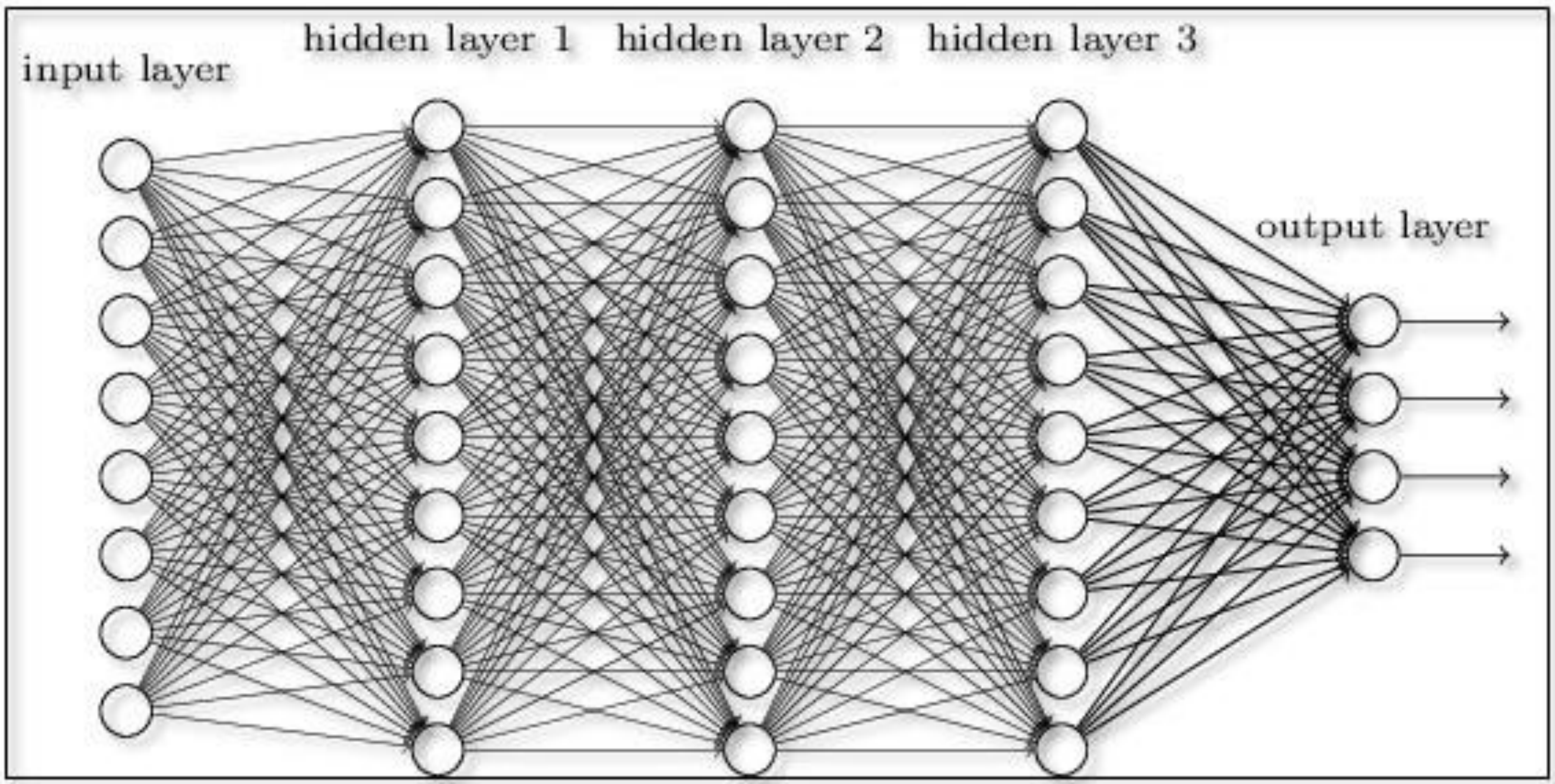
Types of Neurons

- ReLU :Restricted Linear Unit
- F(z) = max(0,z)



Feed forward Neural Netwoks

- Also known multilayer perceptrons
- The foundation of most deep learning models.
- CNNs and RNNs are some special cases of Feedforward networks.
- These networks are called feedforward is that the flow of information takes place in the forward direction


input layer
hidden layer 1
hidden layer 2
hidden layer 3
output layer