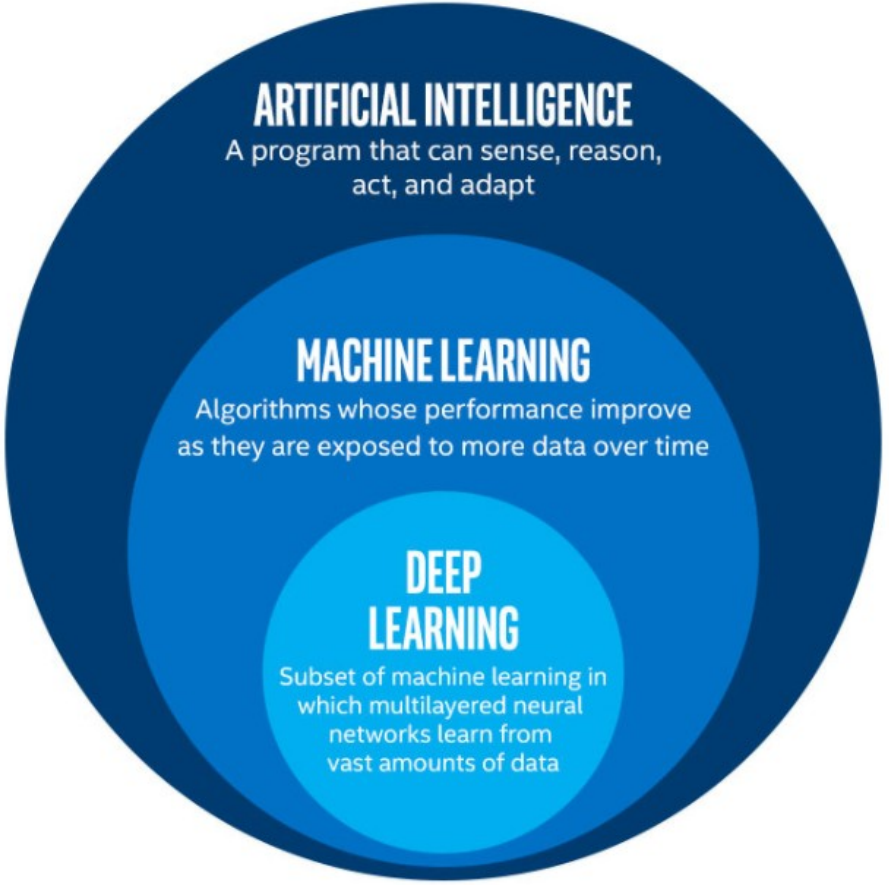
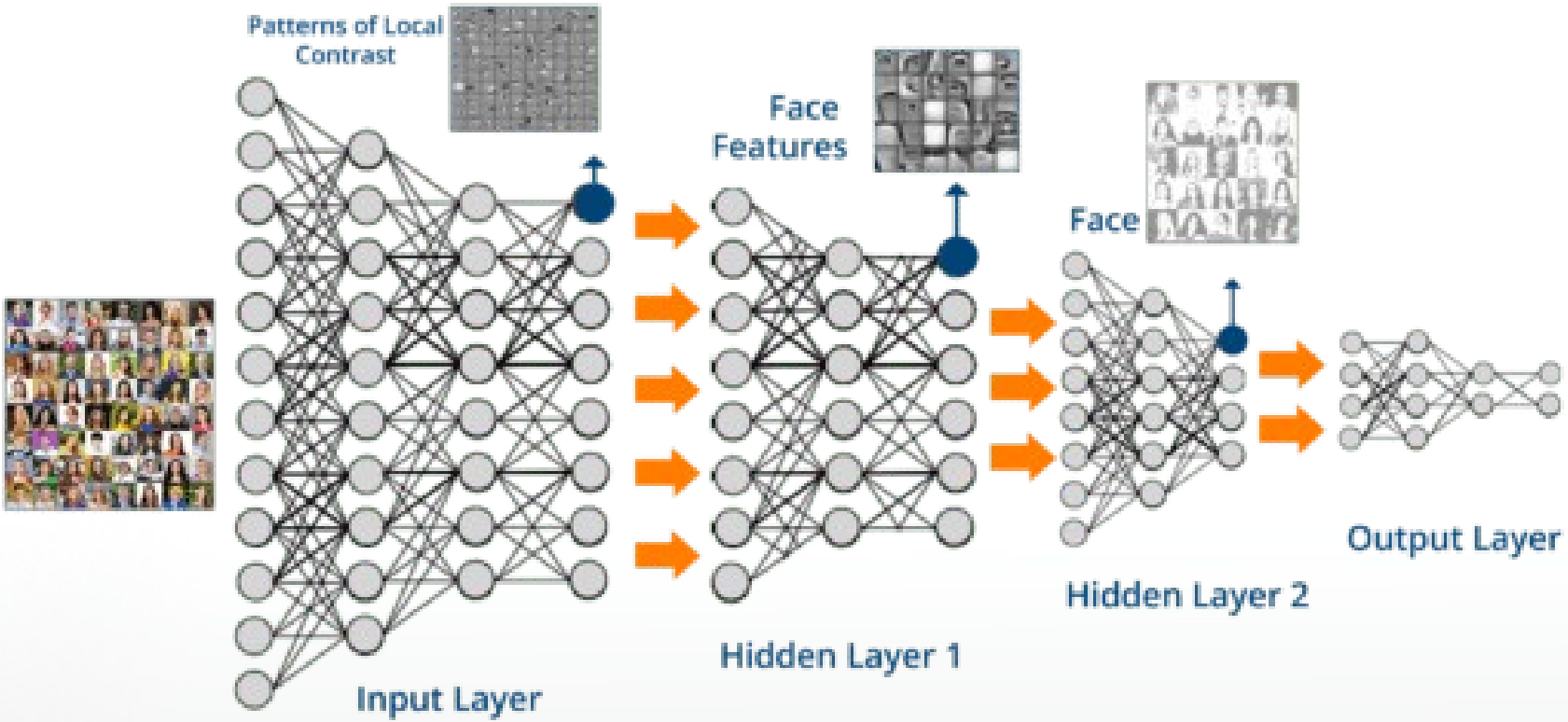
Deep Learning

Deep Learning

- Deep Learning is one of the method by which we can overcome the challange of feature extraction.
- Example :
  - Face regonization system.
  - Object reconization technique
  - Handwritting recognization technique.
  - Discriminate different voices and even recognize person based on his/her voice
  - Natural Langauge Processing

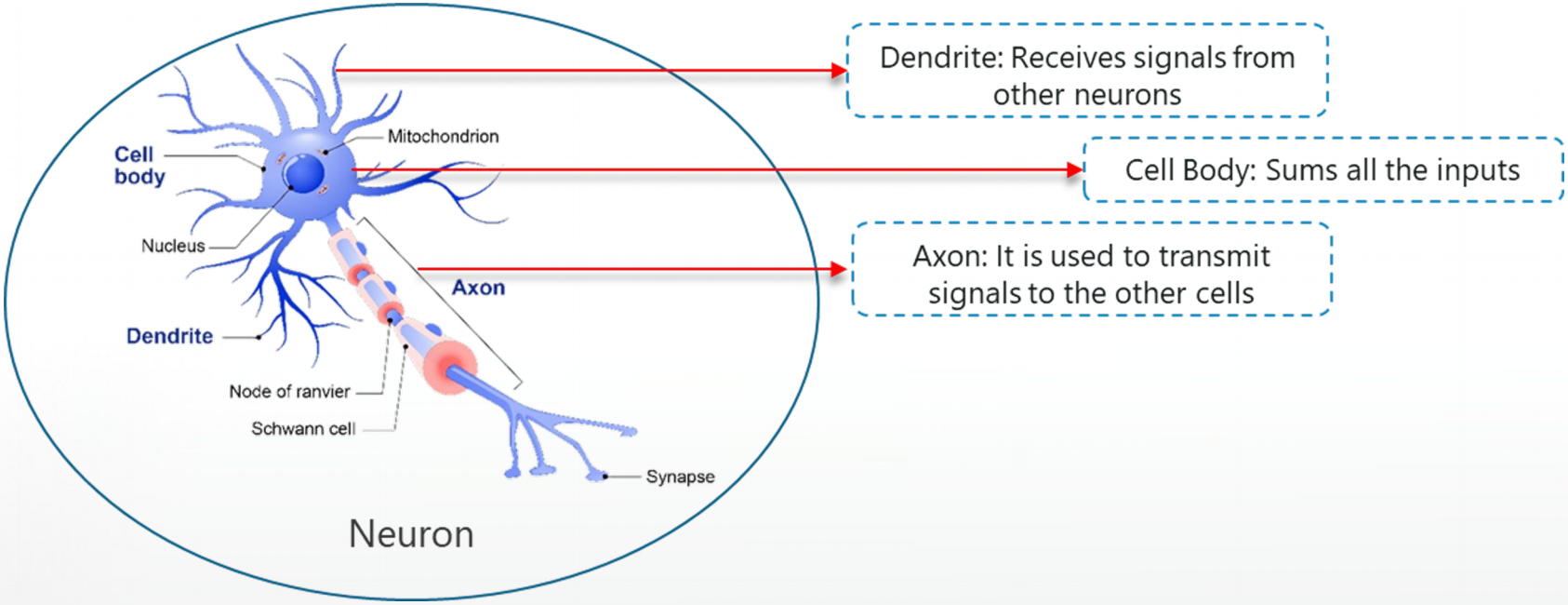


Deep Learning Example



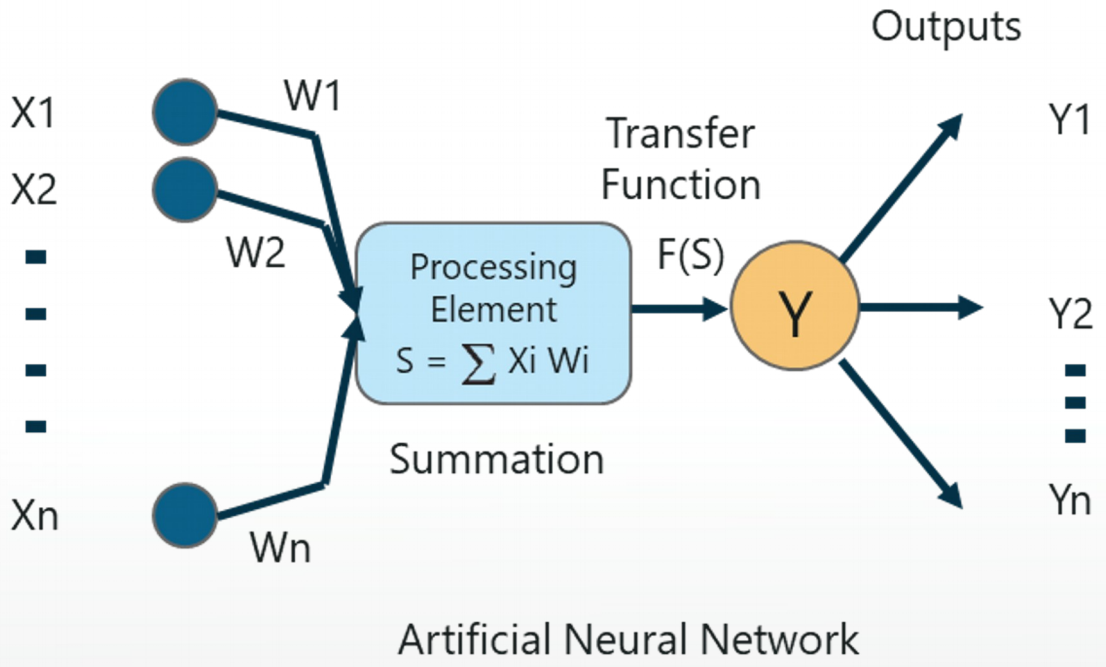
How Deep Learning Works ?

- Deep Learning Studies basic unit of brain called as brain cell or neuron.
- Inspired from nueron an artificial neuron or perceptron was developed.



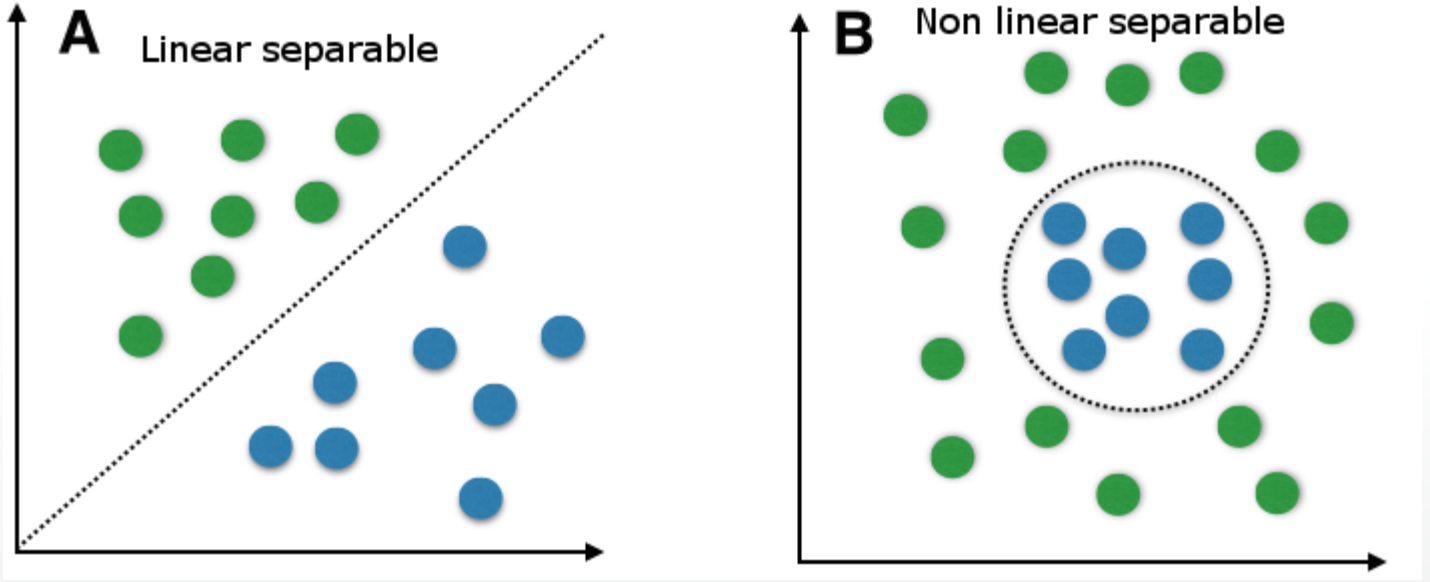
What Is A Perceptron?

- A Perceptron is a single layer neural network that is used to classify linear data or used for binary classification
- Mainly used for supervised learning



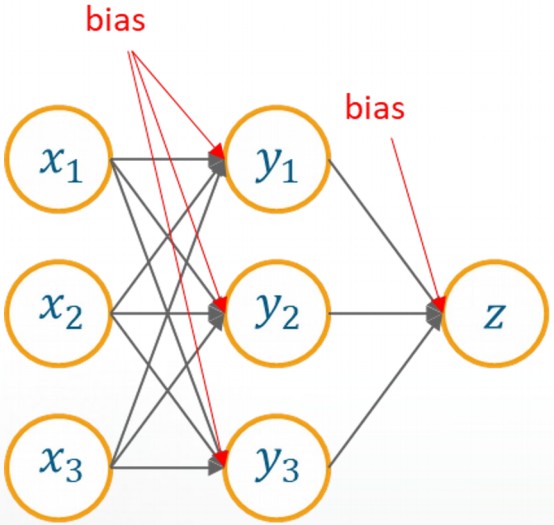
Why Multilayer Perceptron ?

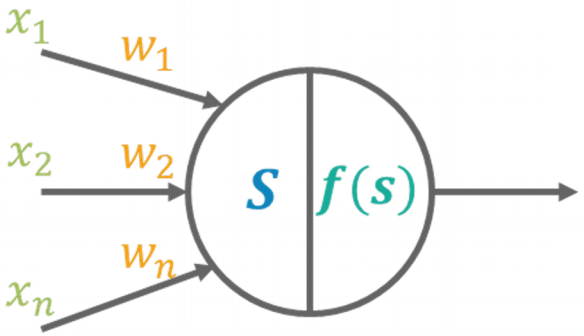
- A Perceptron is a single layer neural network that is used to classify linearly seperable data
- A multilayer perceptron with back propogation is used to resolve this problem



Multilayer Perceptron

- A Multilayer perceptron is a classifier that contains one or more hidden layers.
- It is considered as deep neural network.





Summation:

$$S = \sum_{i=1}^{n} w_i * x_1$$

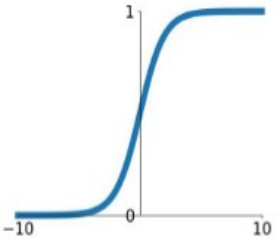
Transformation:

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

Activation Function

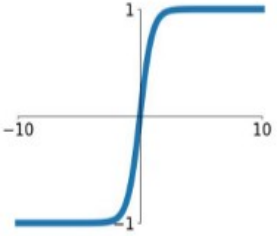
Sigmoid

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



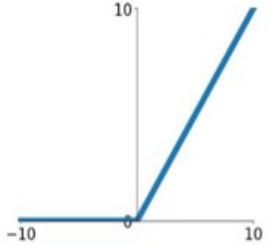
tanh

$$\tanh(x)$$



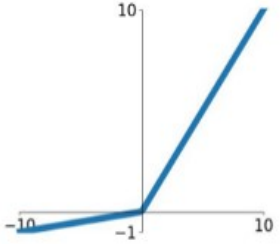
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

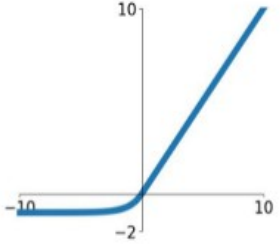


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

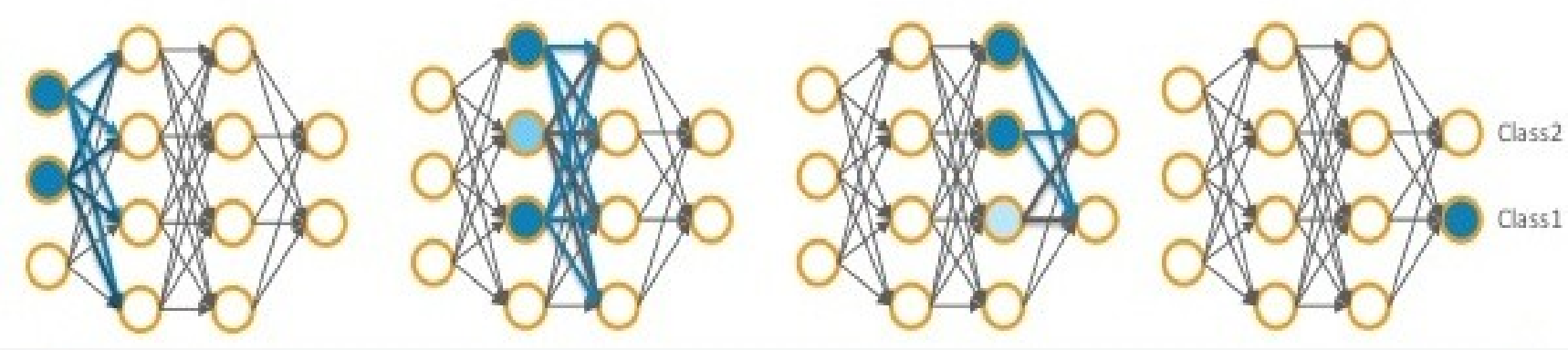
ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

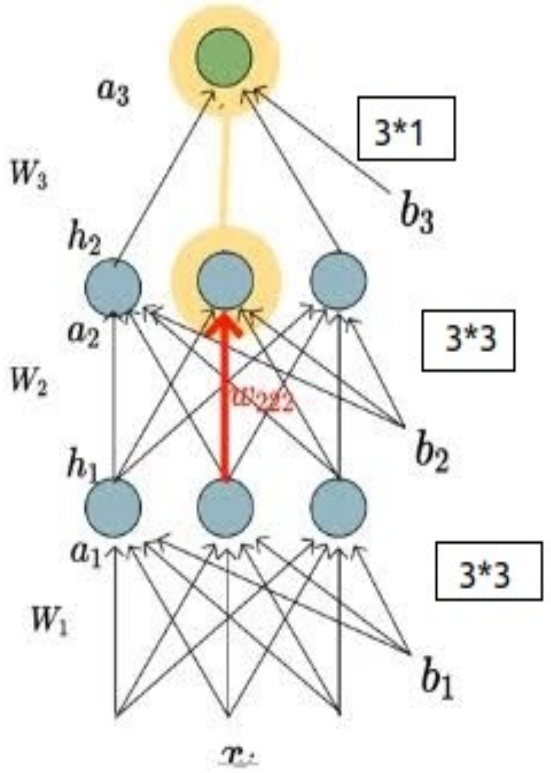


Feed Forward Network

- One of the common deep learning alorihm is backpropogation which is used in feed forward network
- This is used to update weights in shuch a way that most significant variable gets the maximum weight, thus reducing error while compting the output.
- process of updating the weights and training the networks is known as Backpropagation.



Backpropogation



- Let us focus on the highlighted weight (w_{222})
- To learn this weight, we have to compute partial derivative w.r.t loss function

$$(w_{222})_{t+1} = (w_{222})_t - \eta * \left(\frac{\partial L}{\partial w_{222}}\right)$$

learning rate

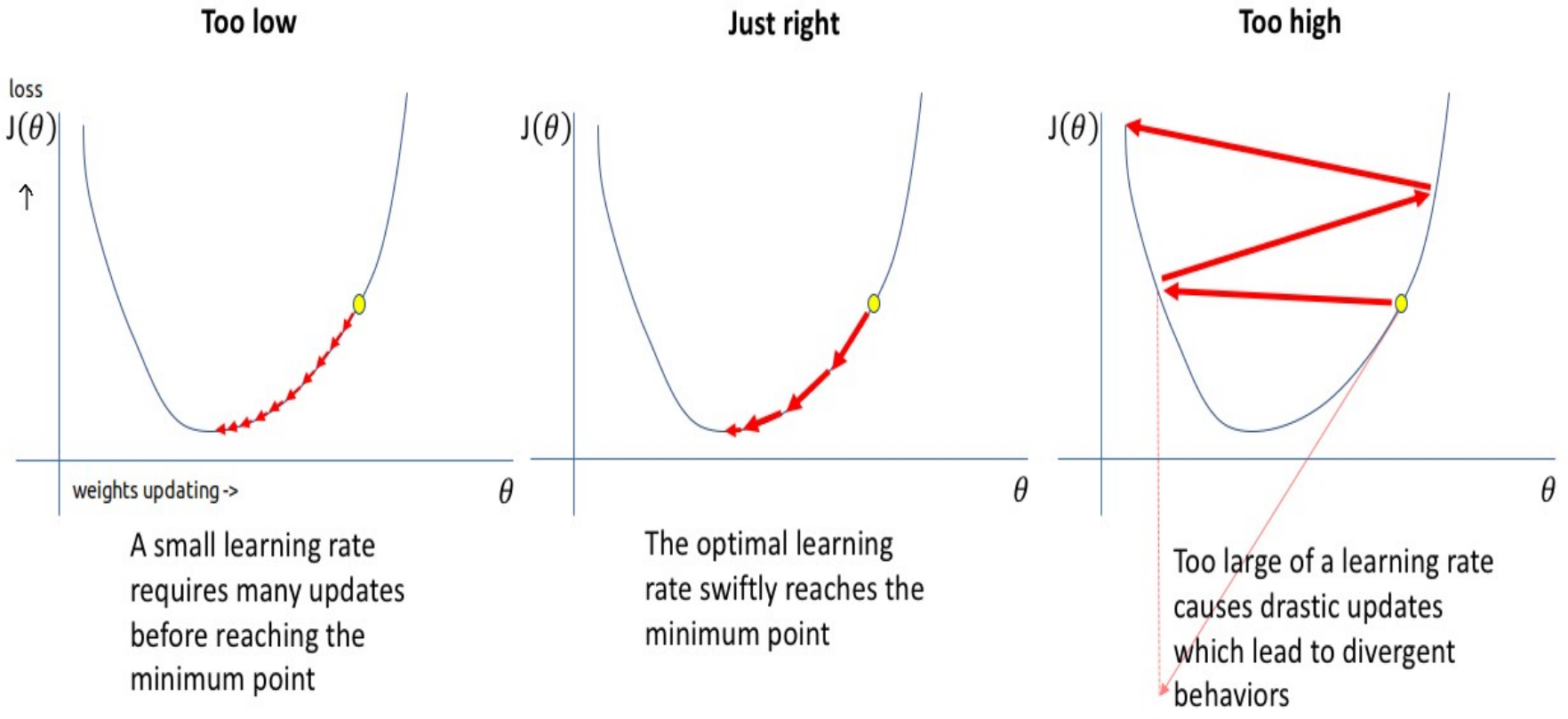
$$\frac{\partial L}{\partial w_{222}} = \left(\frac{\partial L}{\partial a_{22}}\right).\left(\frac{\partial a_{22}}{\partial w_{222}}\right)$$

$$= \left(\frac{\partial L}{\partial h_{22}}\right).\left(\frac{\partial h_{22}}{\partial a_{22}}\right).\left(\frac{\partial a_{22}}{\partial w_{222}}\right)$$

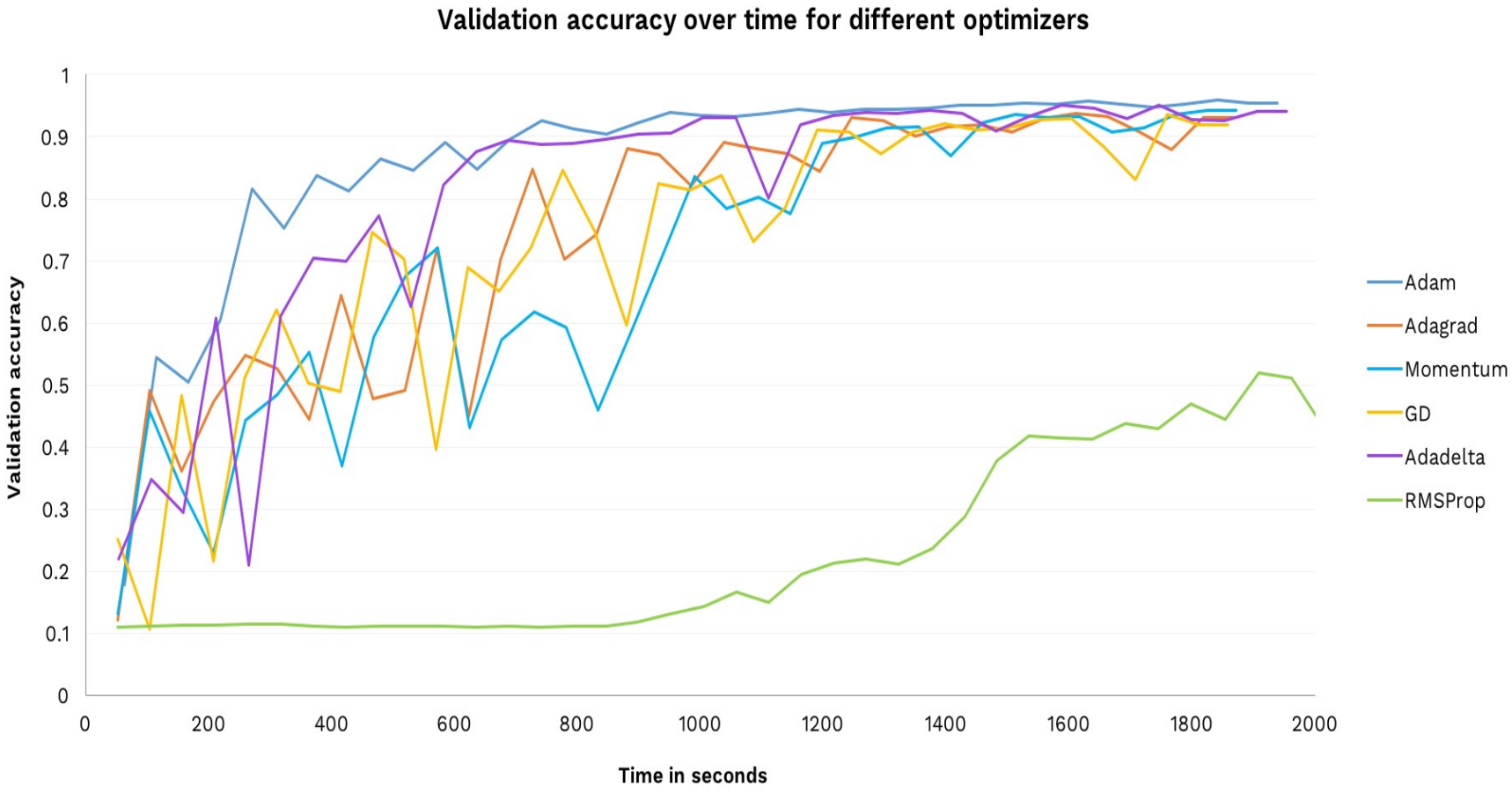
$$= \left(\frac{\partial L}{\partial a_{31}}\right).\left(\frac{\partial a_{31}}{\partial h_{22}}\right).\left(\frac{\partial h_{22}}{\partial a_{22}}\right).\left(\frac{\partial a_{22}}{\partial w_{222}}\right)$$

$$= \left(\frac{\partial L}{\partial \hat{y}}\right).\left(\frac{\partial \hat{y}}{\partial a_{31}}\right).\left(\frac{\partial a_{31}}{\partial h_{22}}\right).\left(\frac{\partial h_{22}}{\partial a_{22}}\right).\left(\frac{\partial a_{22}}{\partial w_{222}}\right)$$

Why Learning rate important ?

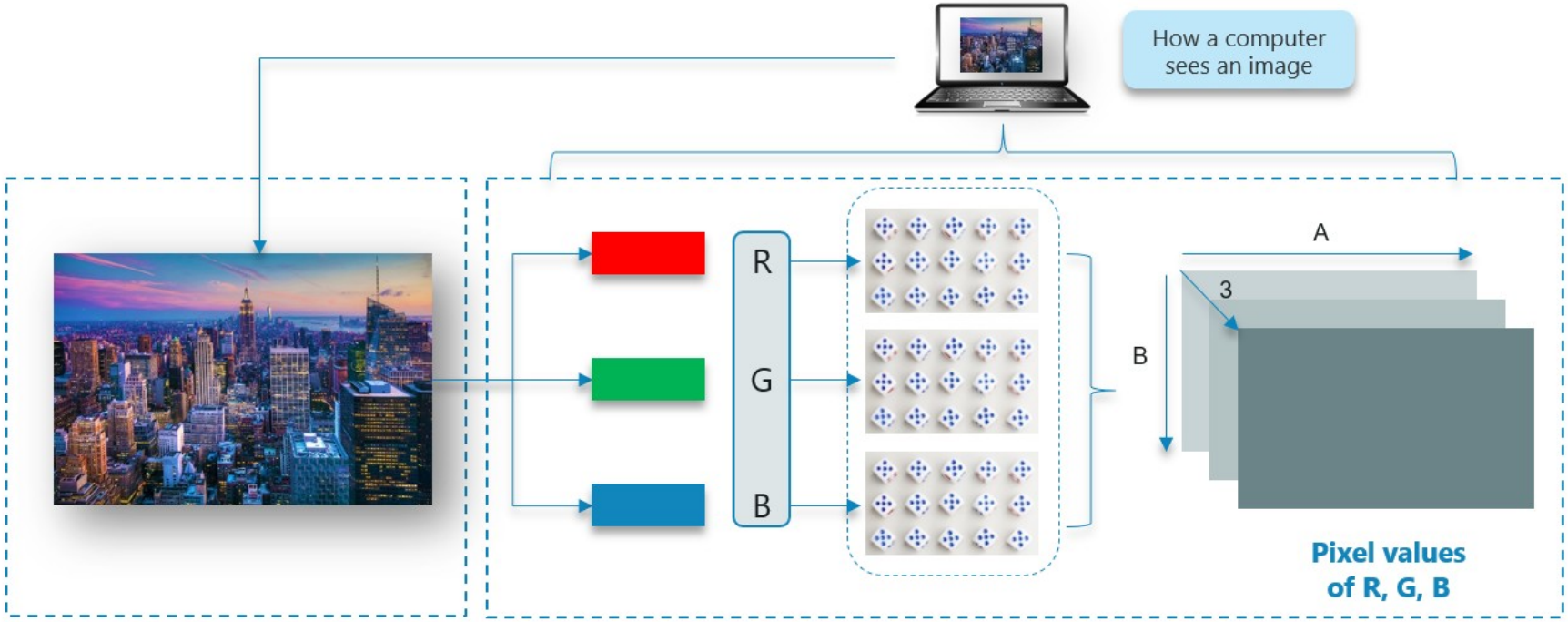


Optimizers


Validation accuracy over time for different optimizers
Validation accuracy
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Time in seconds
Adam
Adagrad
Momentum
GD
Adadelta
RMSProp

Convolutional Neural Network

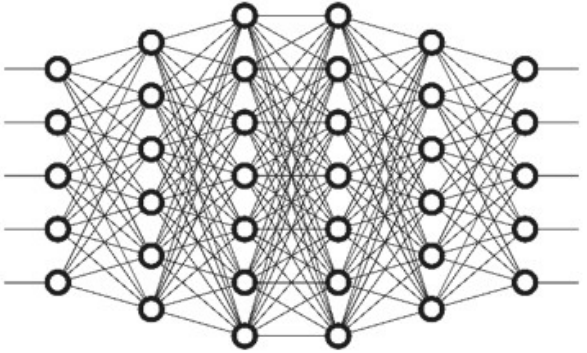
- Convolutional Neural Networks are designed to address image recognition systems and classification problems.



Why Convolutional Neural Network ?

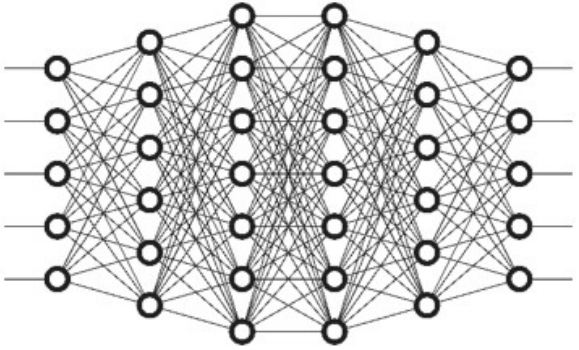
- We cannot make use of fully connected networks when it comes to Convolutional Neural Networks.

Image with 28 x 28 x 3 pixels



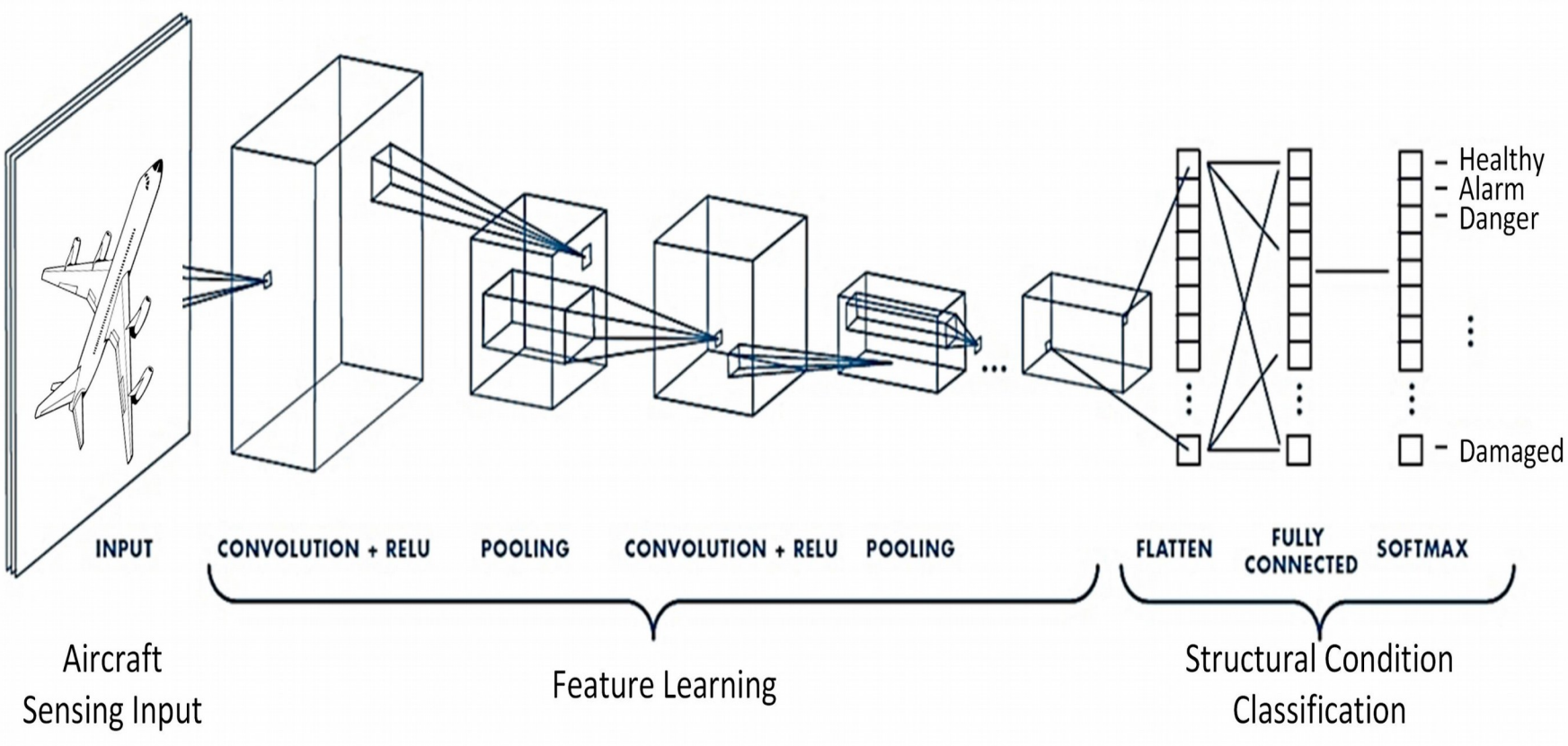
Number of weights in the first hidden layer will be 2352

Image with 200 x 200 x 3 pixels

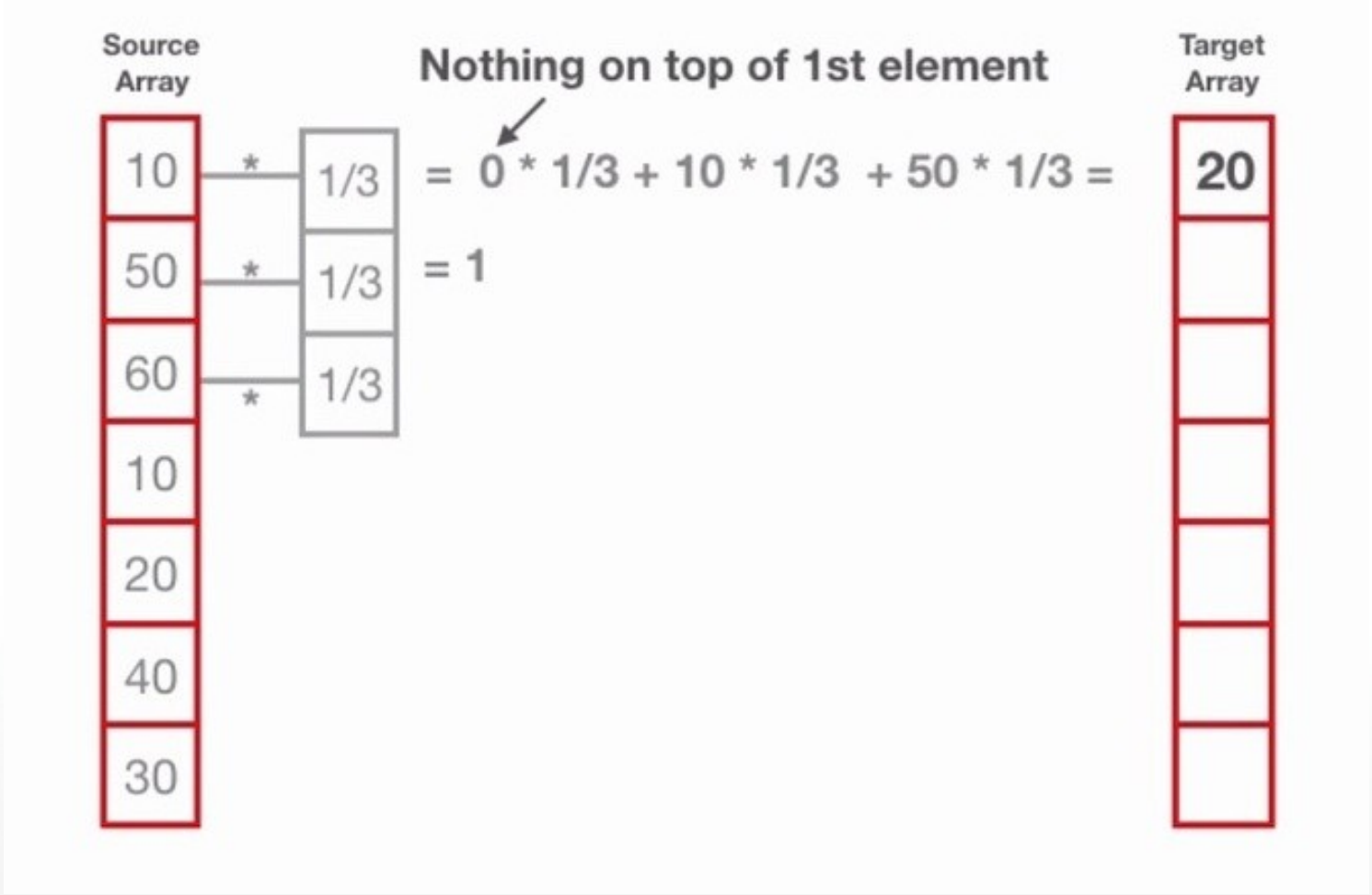


Number of weights in the first hidden layer will be 120,000

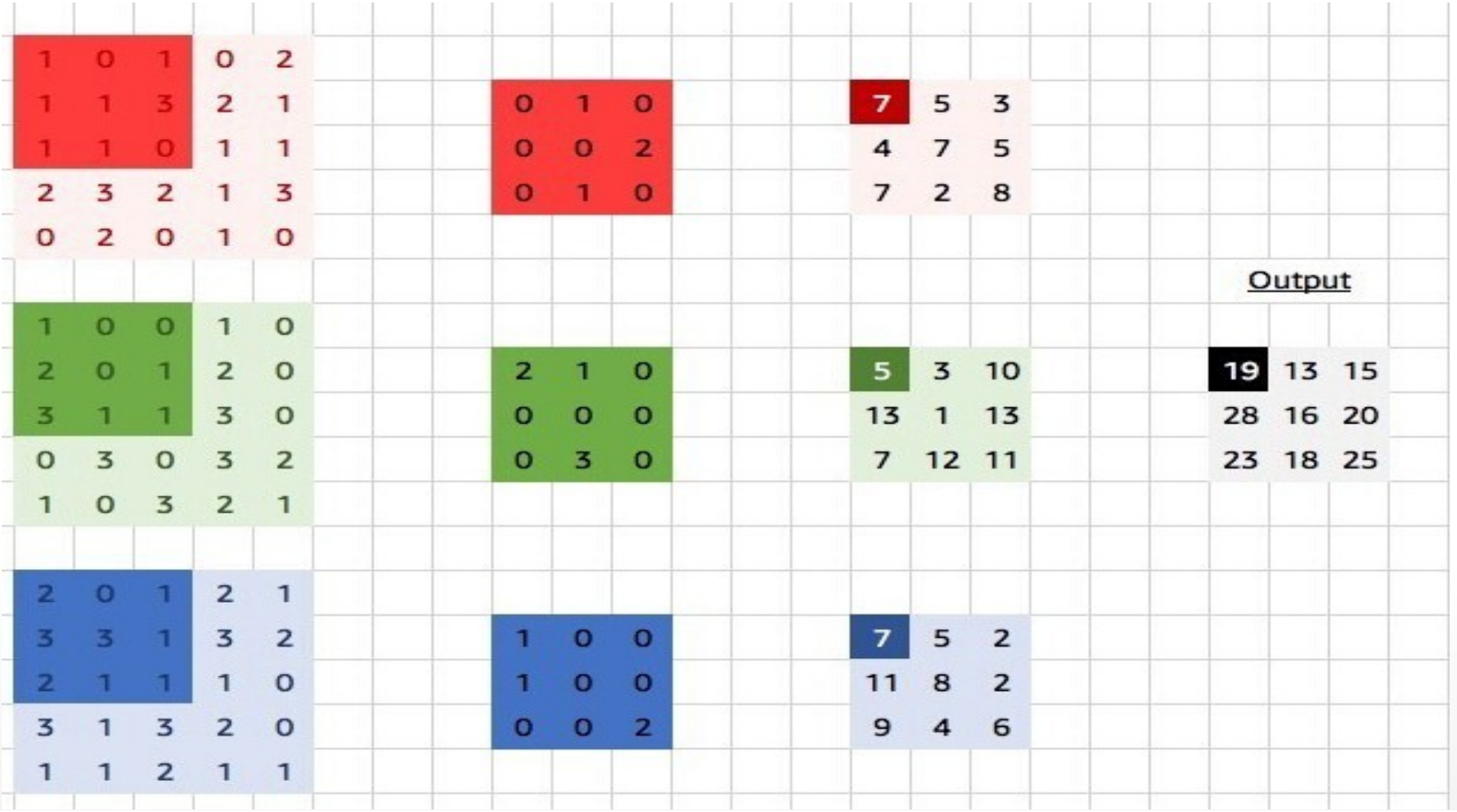
How CNN works ?



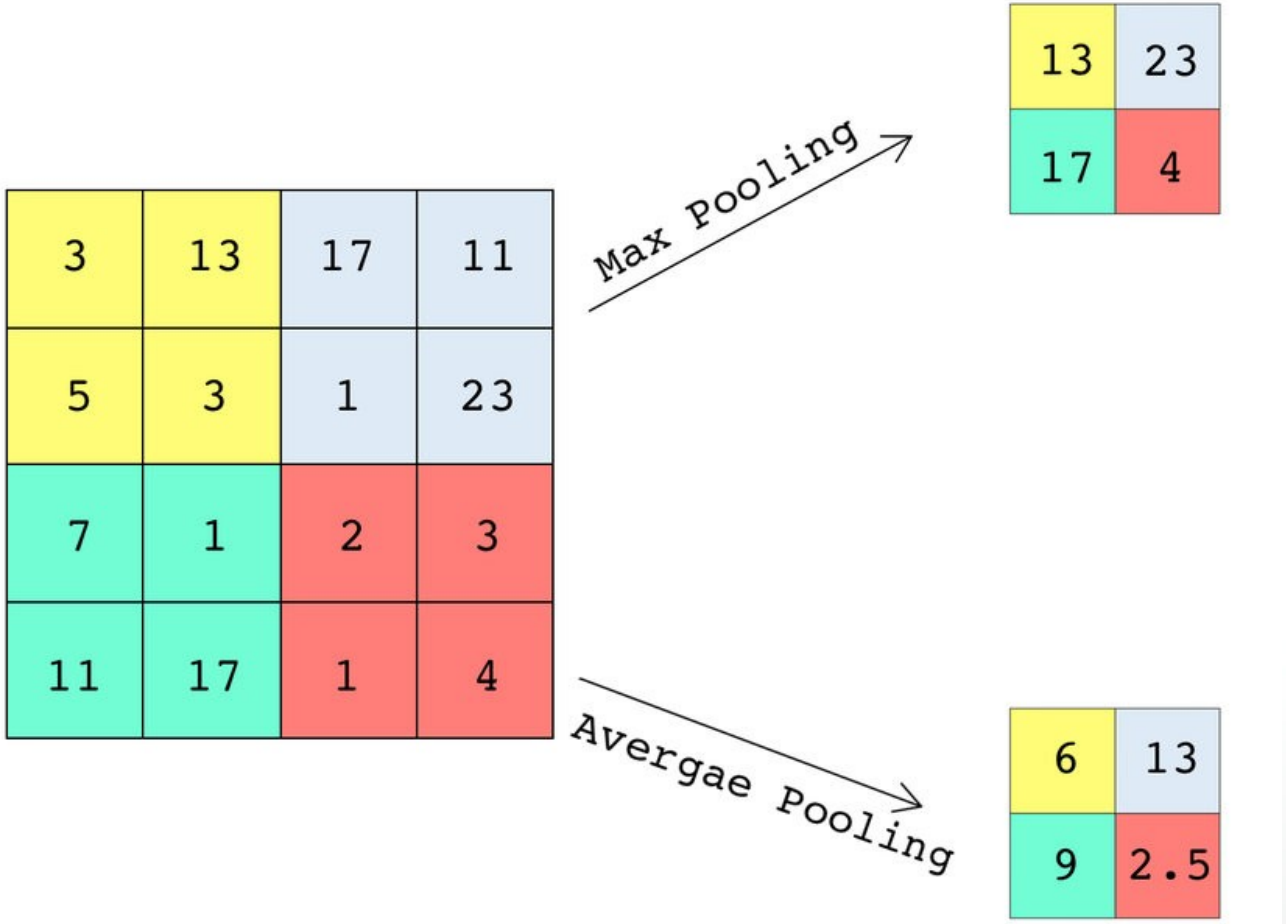
1D Convolution



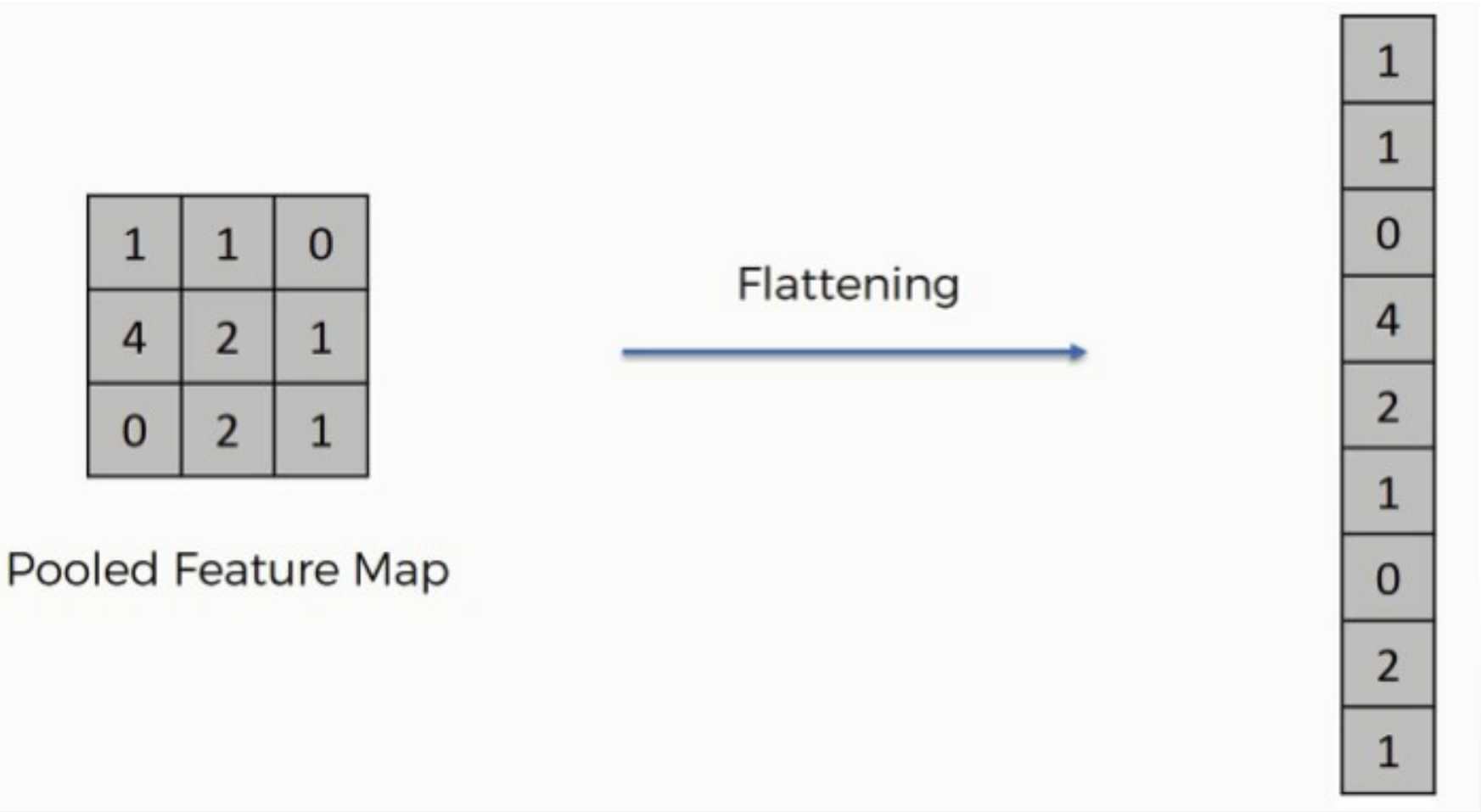
2D Convolution



Pooling


3
13
17
11
5
3
1
23
7
1
2
3
11
17
1
4
Max Pooling
13
23
17
4
Avergae Pooling
6
13
9
2.5

Flatten



Recurrent Neural Network

- Feed-forward networks cannot be used when predicting a word in a sentence as it will have no absolute relation with the previous set of words.
- But, with Recurrent Neural Networks, this challenge can be overcome.