Part 3
Lecture 1b: Neural Networks

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

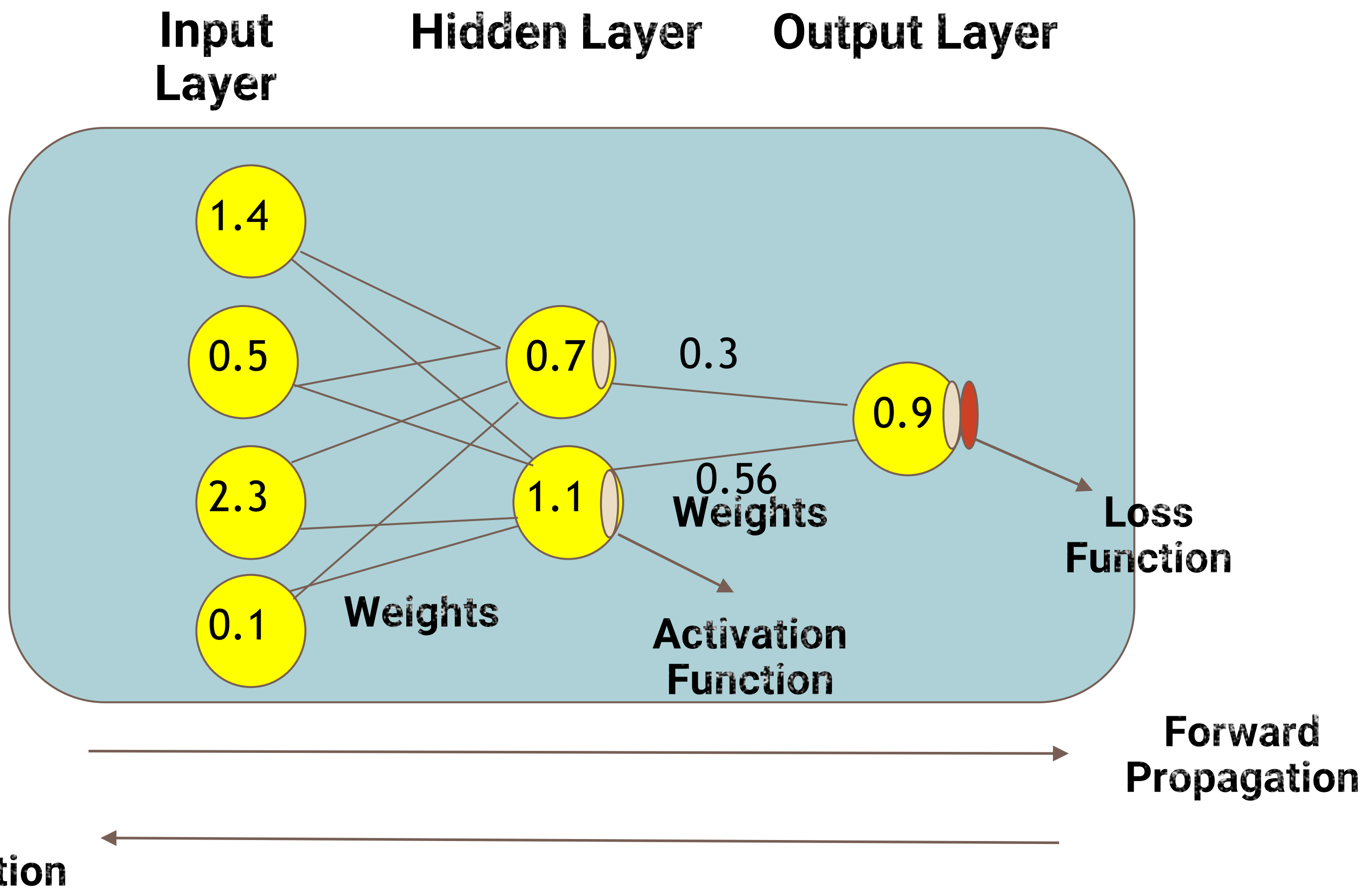
Who I am…

Pascal Tyrrell, PhD *Associate Professor*

Department of Medical Imaging, Faculty of Medicine

Institute of Medical Science, Faculty of Medicine
Department of Statistical Sciences, Faculty of Arts and Science

Neural Networks

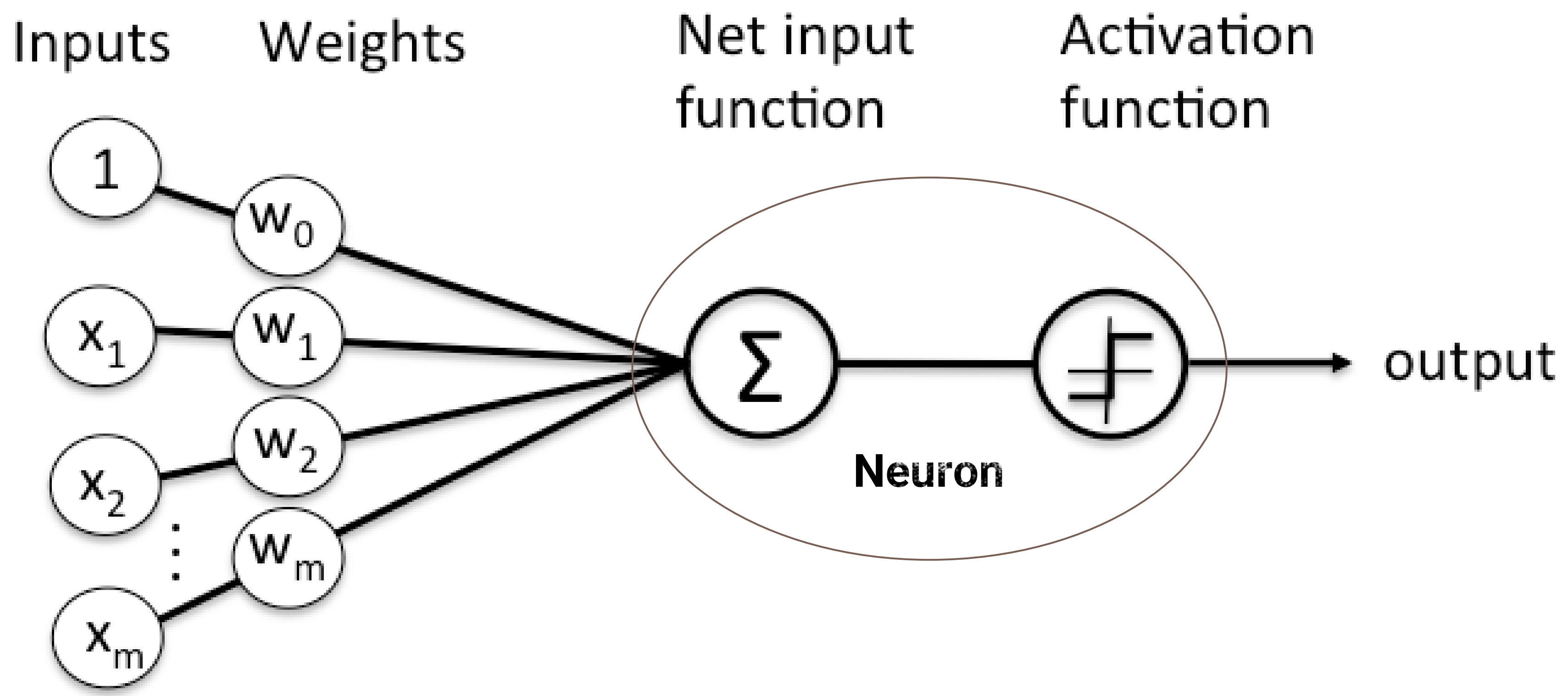


Epochs:
Iterations, going forward and then back propagation counts as one.

Batch size:
Separate data into batches to lighten the load on the GPU, one epoch is counted when all batches have completed.

Layers of neurons with linear combinations resulting in the neuron value, the activation function and weights.

Neuron

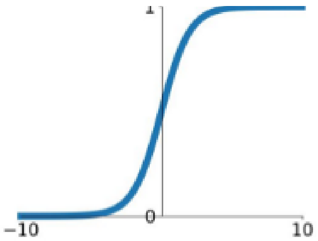


A neuron is made up of the linear combination of the inputs and the weights that later couples with the activation function.

Activation Functions: middle layers

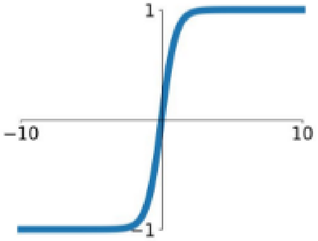
Sigmoid

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



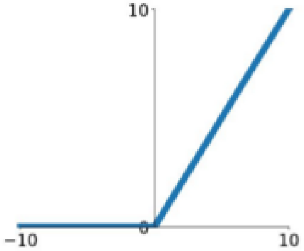
tanh

$\tanh(x)$



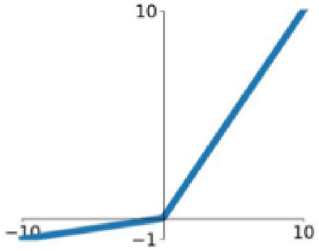
ReLU

$\max(0, x)$



Leaky ReLU

$\max(0.1x, x)$

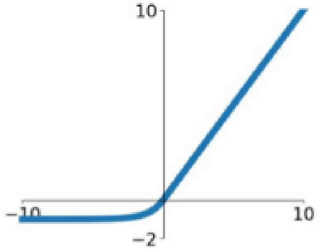


Maxout

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

ELU

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



The activation function reflects the behavior of the neurons with some specific inputs.

Activation Functions: Output Layer

Regression:

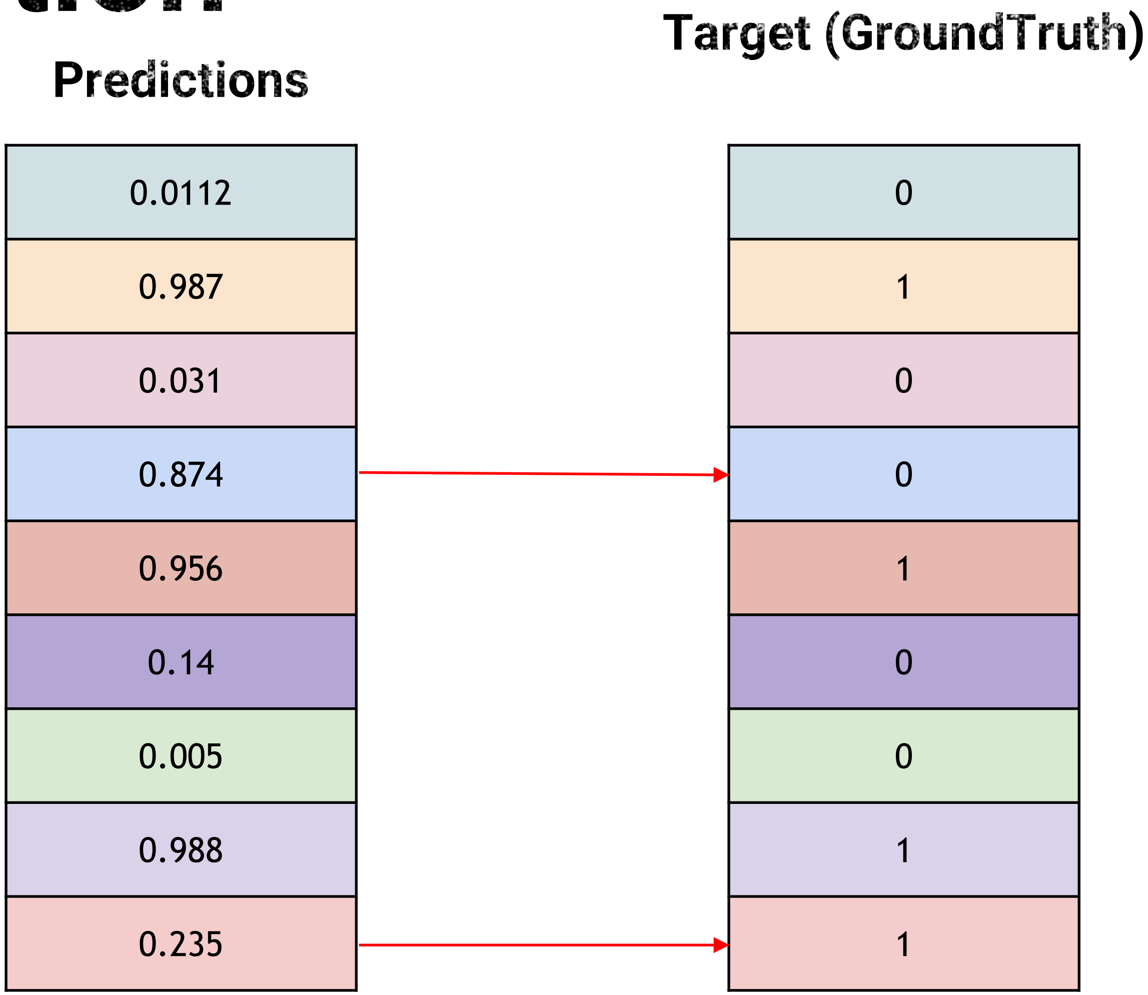
1. Linear
2. Tansig

Classification:

1. Softmax
2. Sigmoid
3. Tanh

Any function can be used as activation function.
Guide: https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/

Loss Function

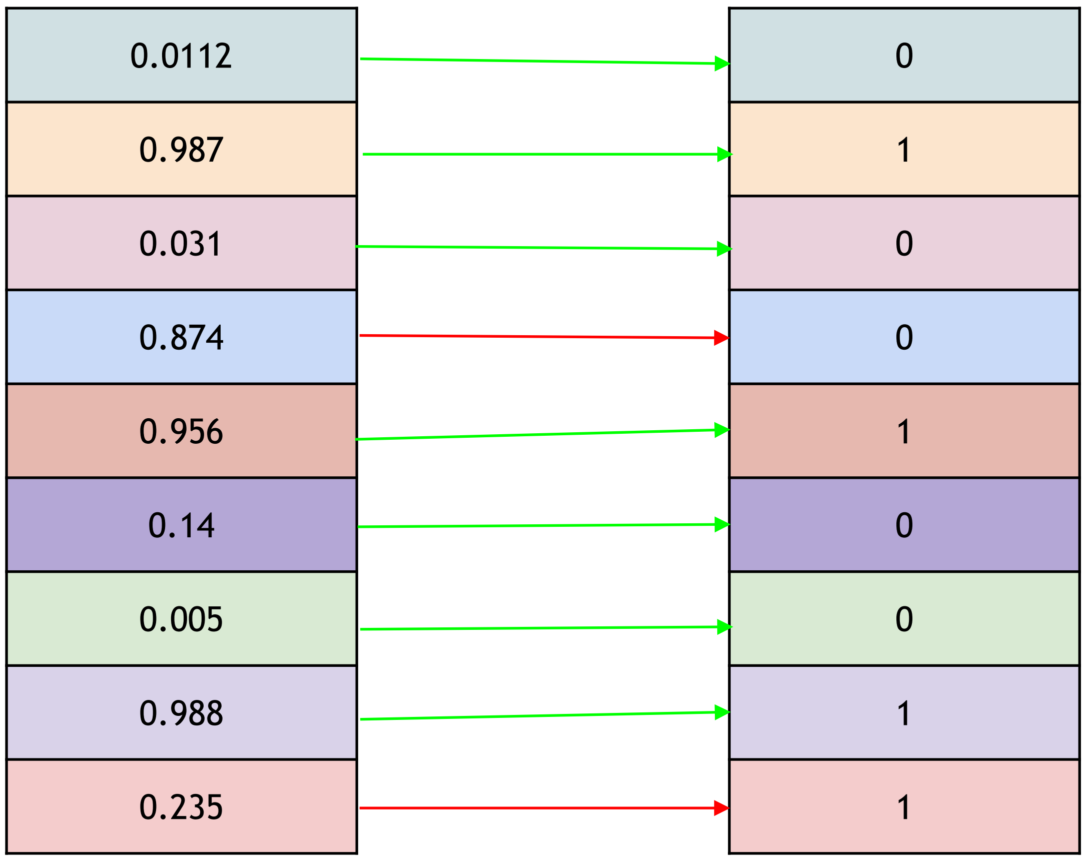


Predictions	Target (GroundTruth)
0.0112	0
0.987	1
0.031	0
0.874	0
0.956	1
0.14	0
0.005	0
0.988	1
0.235	1

Error:
Differences between the prediction and the target

Neural Networks are universal approximators. They try to imitate the function of the dataset.

Loss Function



Predictions	Target (GroundTruth)
0.0112	0
0.987	1
0.031	0
0.874	0
0.956	1
0.14	0
0.005	0
0.988	1
0.235	1

Very Basic Loss Function:

Round the predictions and sum the #correct over the total#

Error: 2/9

The loss function is used to calculate the error.

Recommended Loss Functions

Regression:

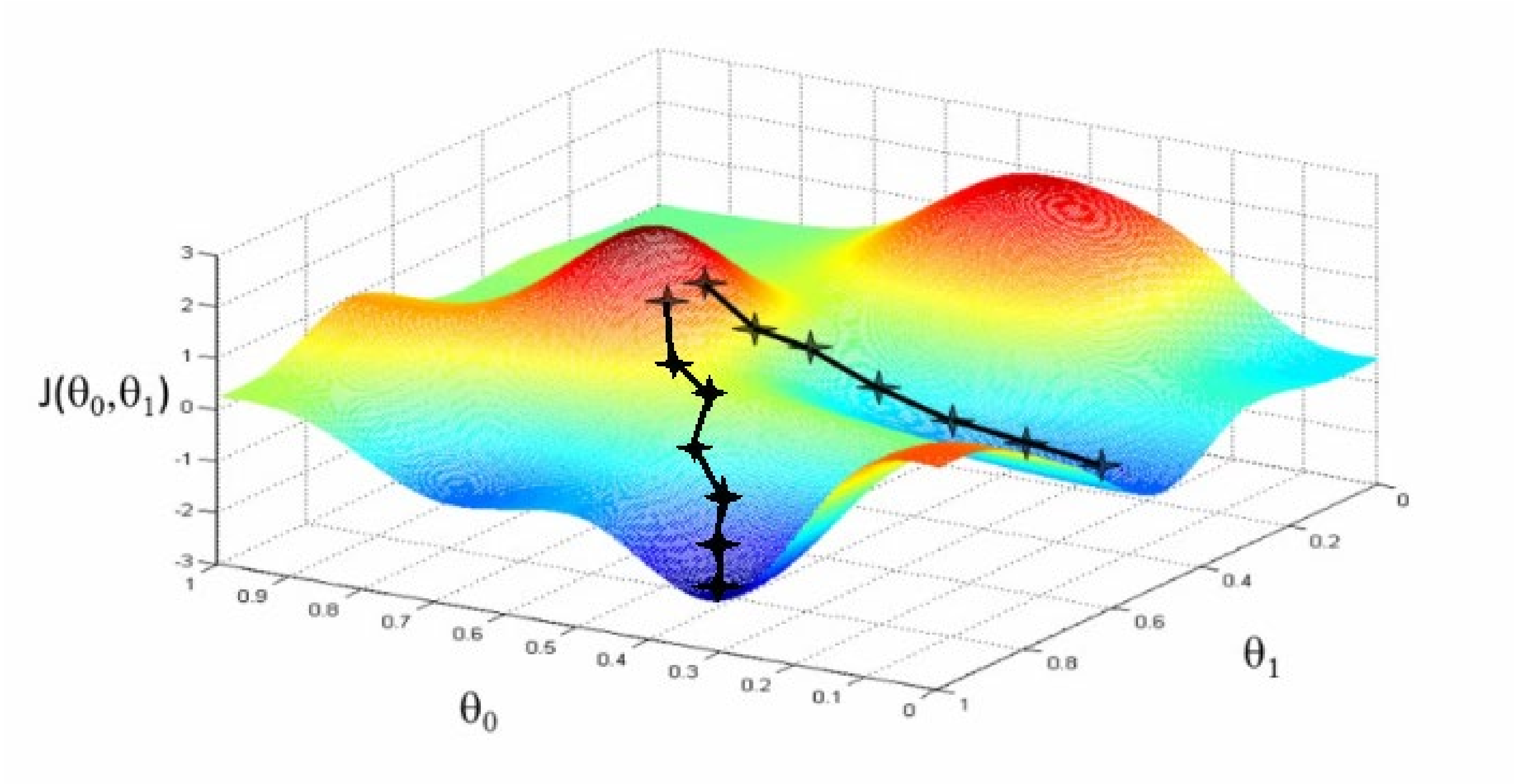
1. Mean Squared Error: More penalization to outliers
2. Mean Absolute Error: More robust to outliers

Classification:

1. Cross Entropy: Multiclass classification
2. Binary Cross Entropy: Binary classification
3. Hinge: Penalization for uncertain predictions

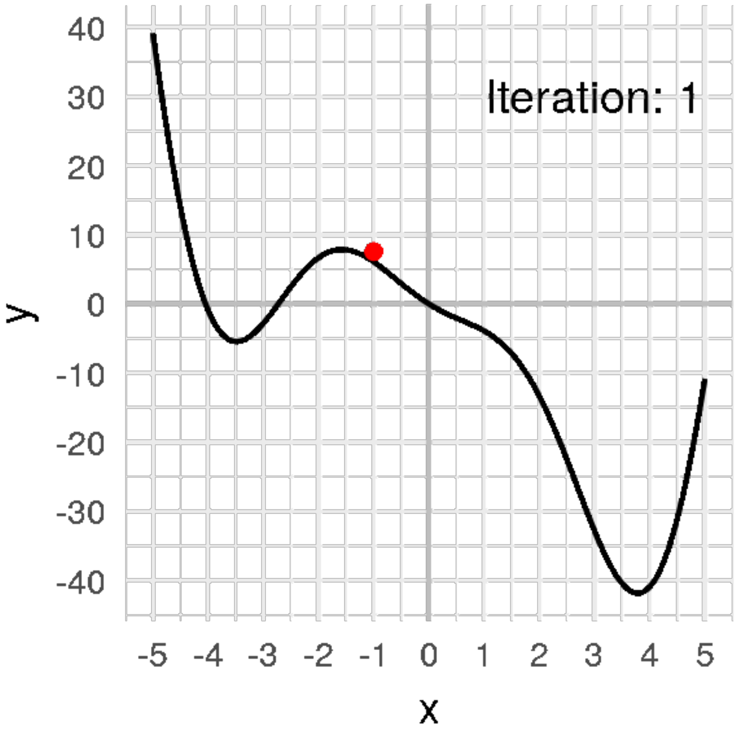
Any function can be used as loss function.
Guide: https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23

Optimizer

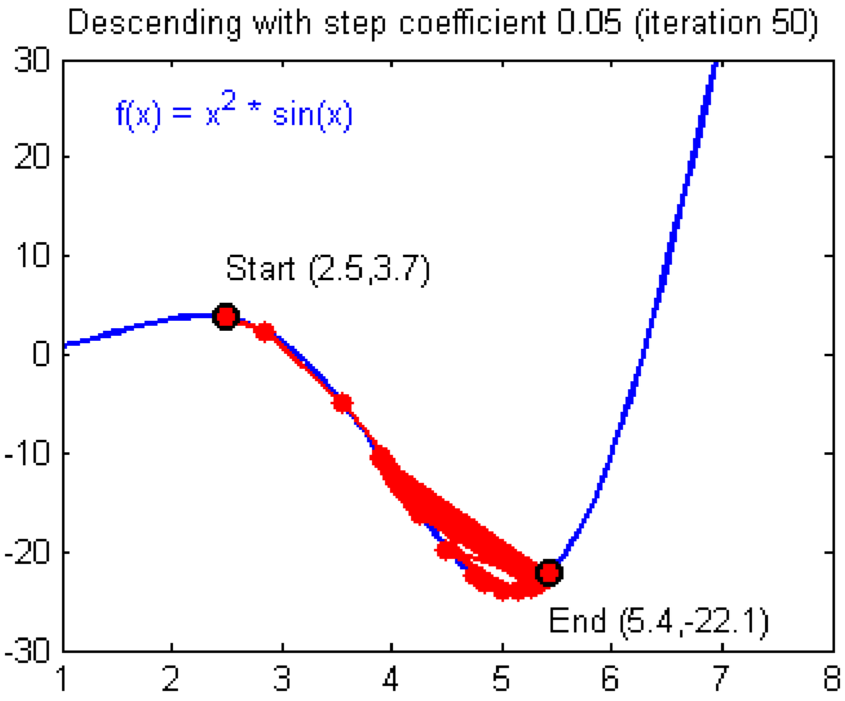


The loss function is a surface: we are trying to find our minima for best performance

Optimizer - Learning rate



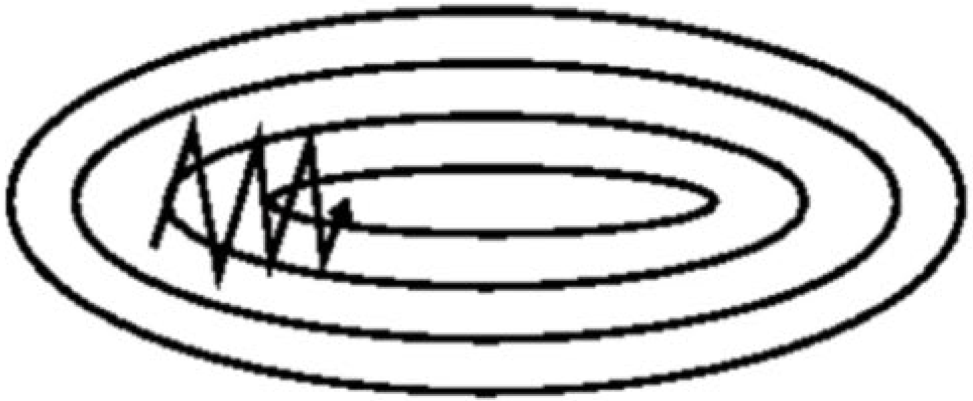
Small



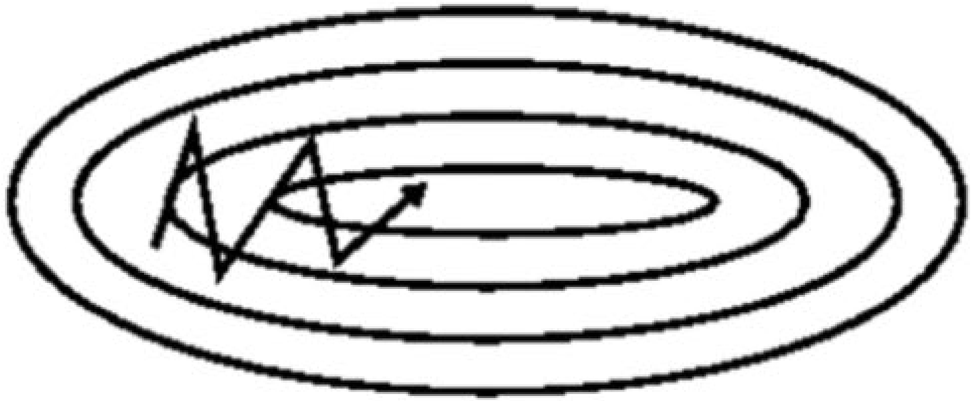
Big

Small learning rate: Slow, may not find the minima
Big learning rate: Fast, may jump over the minima
Solution: Adaptive gradient

Optimizer - Momentum



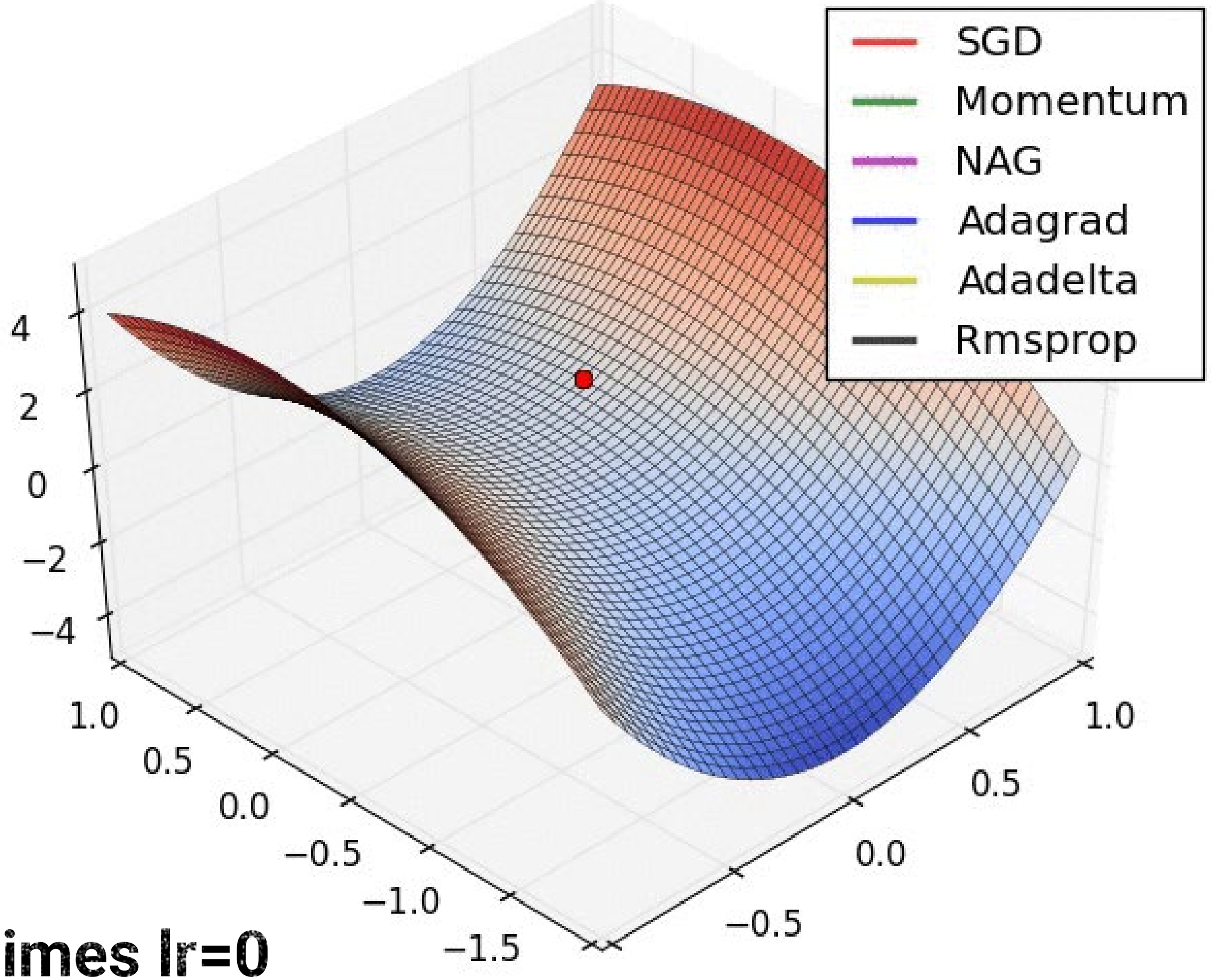
No momentum



With Momentum

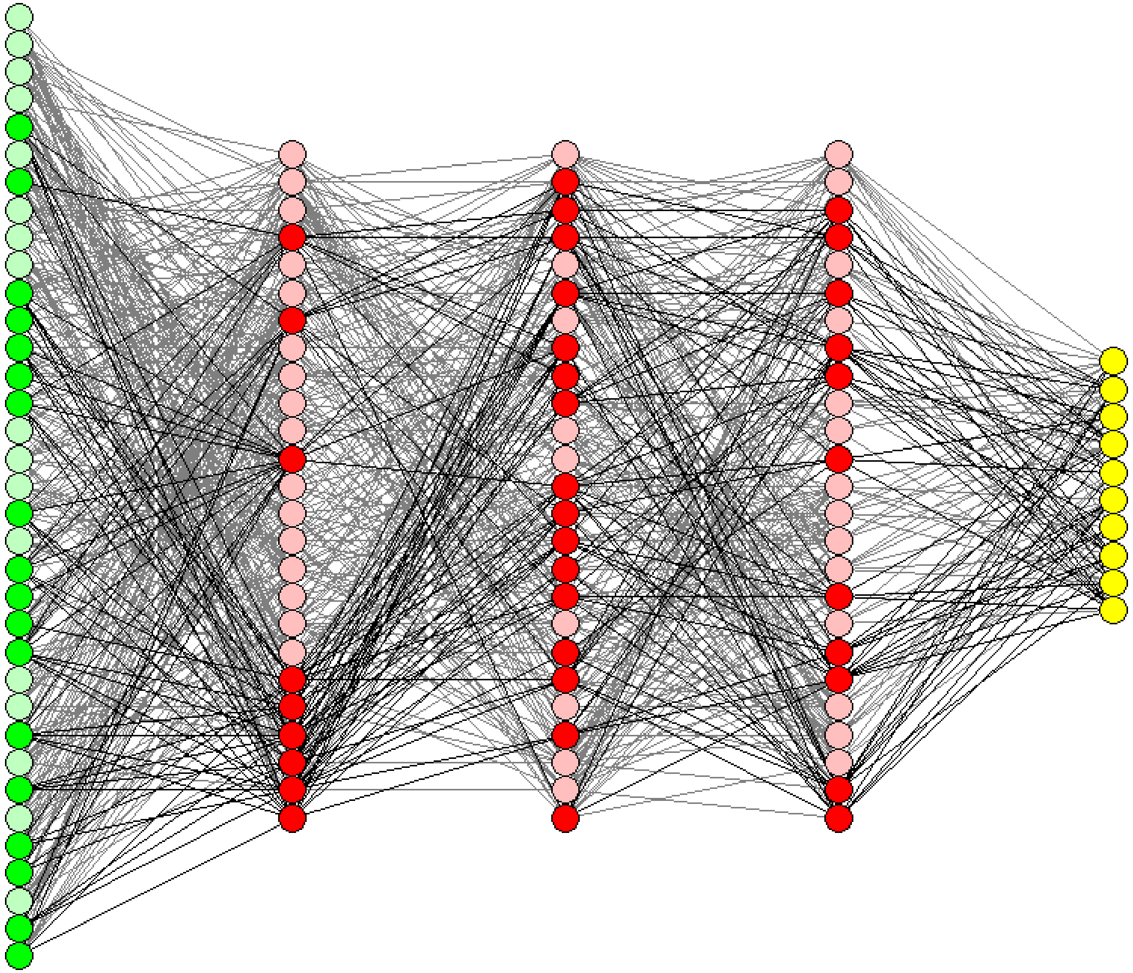
Reduce the amount of steps to get to the minima.
Sometimes the momentum can be so big that it passes through the minima.
Solution: Adaptive momentum

Optimezers



1. **Stochastic Gradient Descent (SGD)**
2. **Momentum**
3. **Adaptive Gradient (AdaGrad): sometimes lr=0**
4. **Adaptive Delta (AdaDelta)**
5. **Adaptive Momentum (Adam): Recommended**

Deep Learning



Adding more layers we increase the possible number of approximations but we are adding complexity.

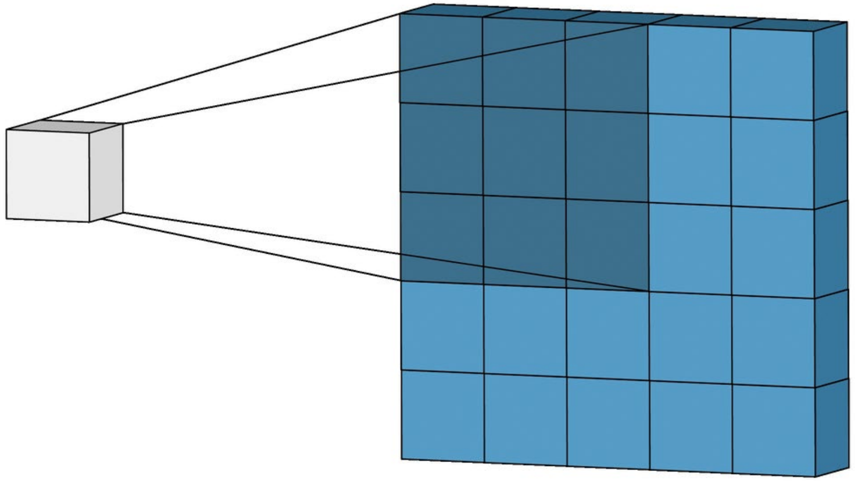
DL: What does it do?

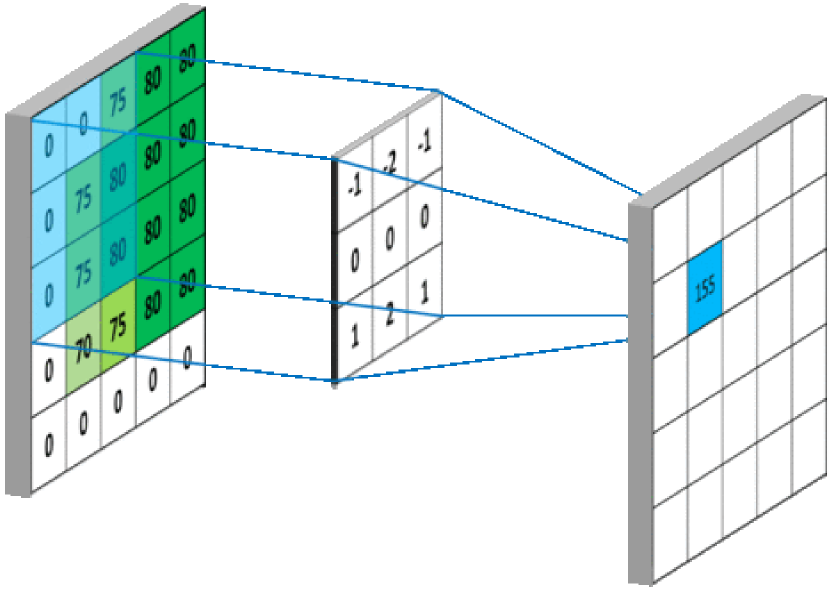
- ❑ Deep learning algorithms can automatically learn feature representations from data without the need for prior definition by experts.

Types of Layers in the CNN

1. Input
2. Convolutional
3. Pooling
4. Batch Normalization
5. Dropout
6. Activation Function (ReLU, Softmax or other)
7. Flattening
8. Fully Connected

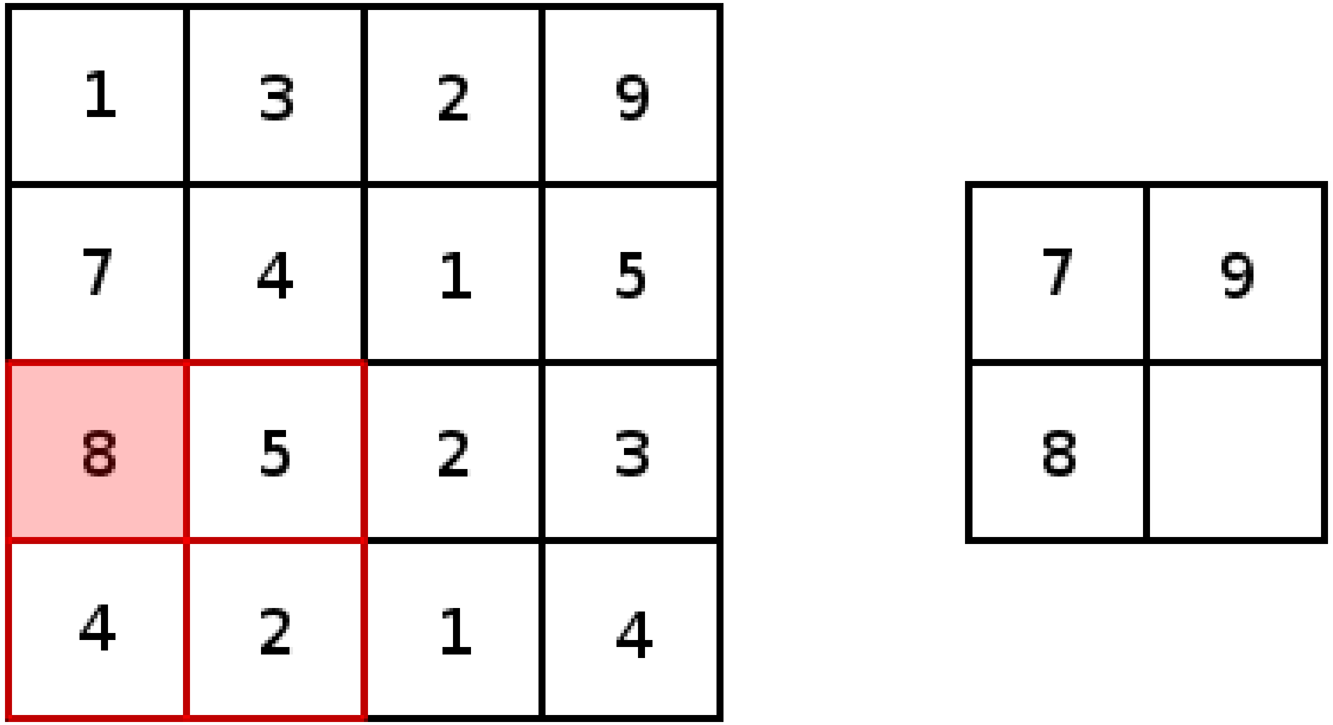
Convolutional Layer





Main component of Convolutional Neural Networks.
It learns many filters (a set of learnable weights) to extract characteristics from the images. You can think of a filter as storing a single template or pattern.

Pooling Layer



1	3	2	9
7	4	1	5
8	5	2	3
4	2	1	4

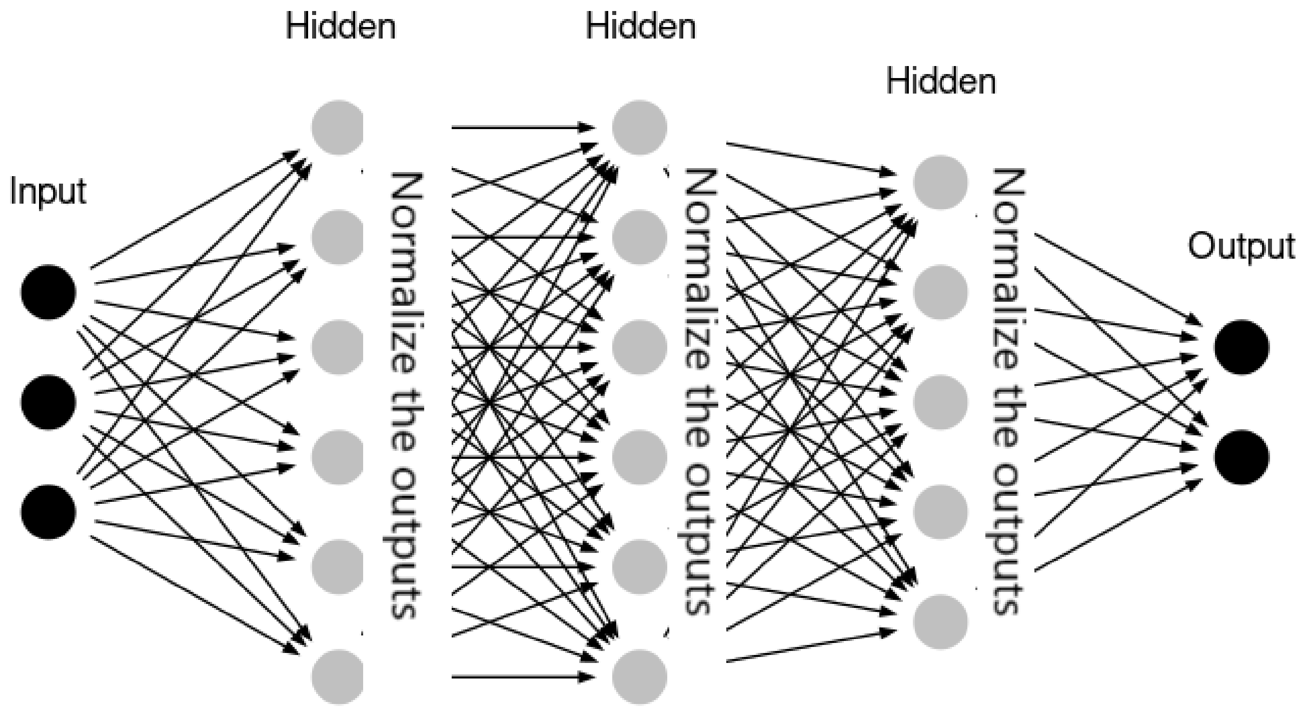
7	9
8	

The Convolutional Layer extracts to much information.
Pooling Layer reduces redundant or insignificant features. Reduce space and computation time.
There are several ways to do pooling, e.g. max pooling and average pooling

Medical Imaging
UNIVERSITY OF TORONTO

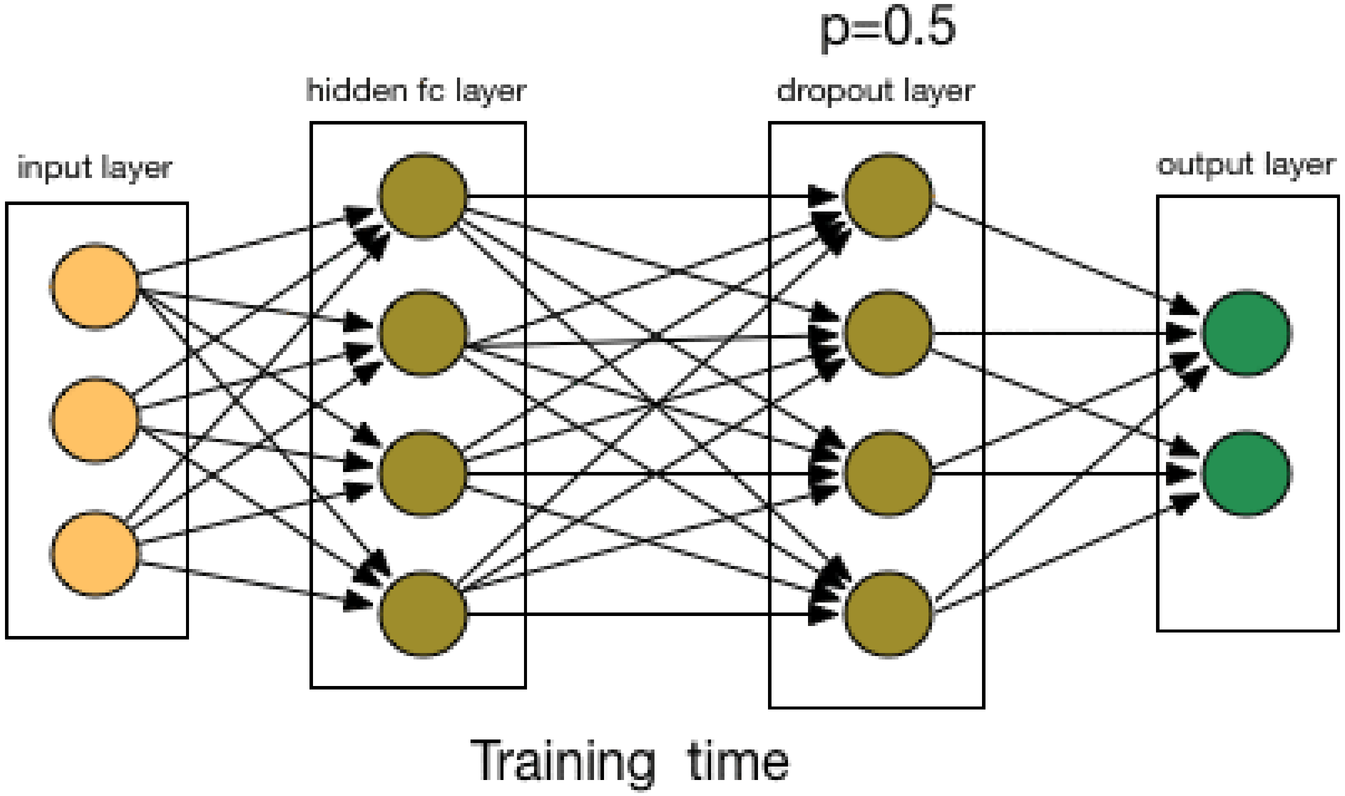
MiDATA

Batch Normalization Layer



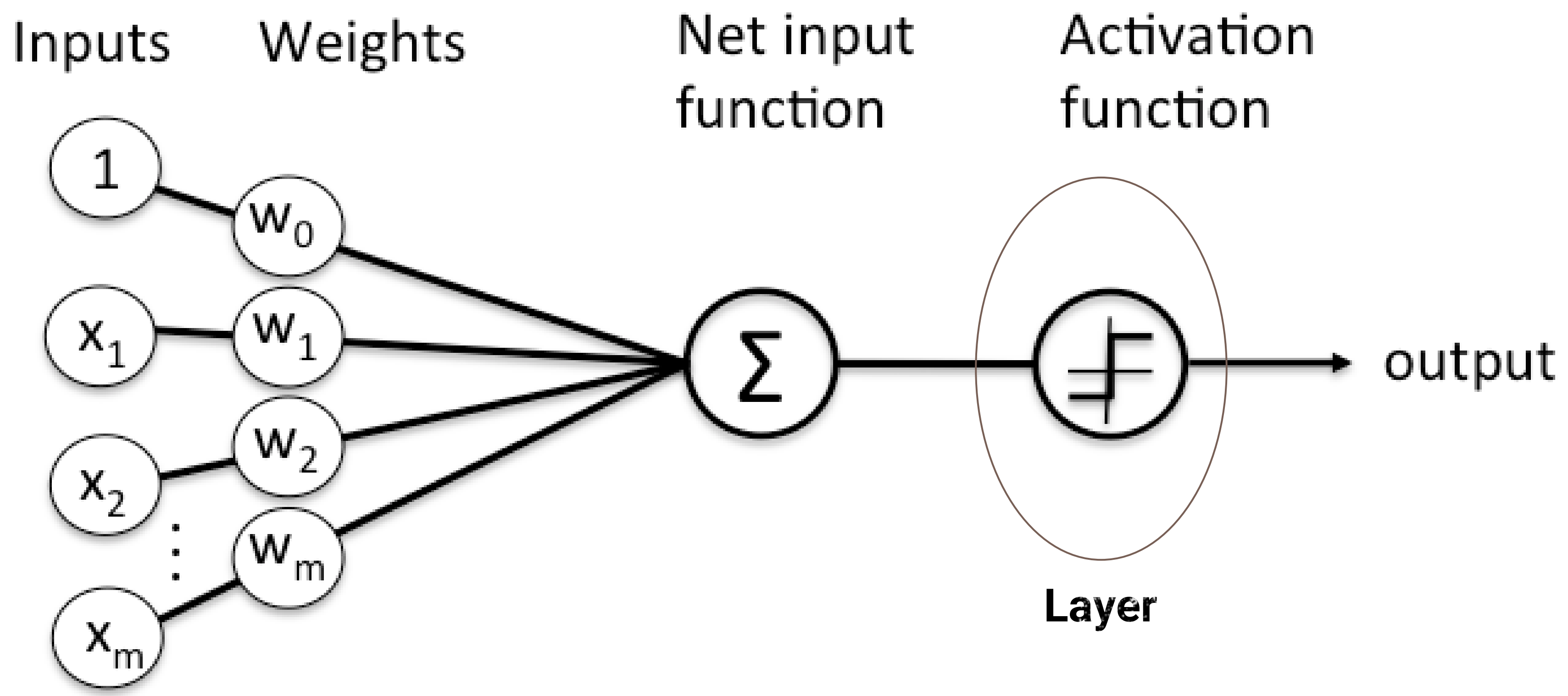
Normalize every output to the same scale.
Allows the network to learn better and more easily.

Dropout Layer



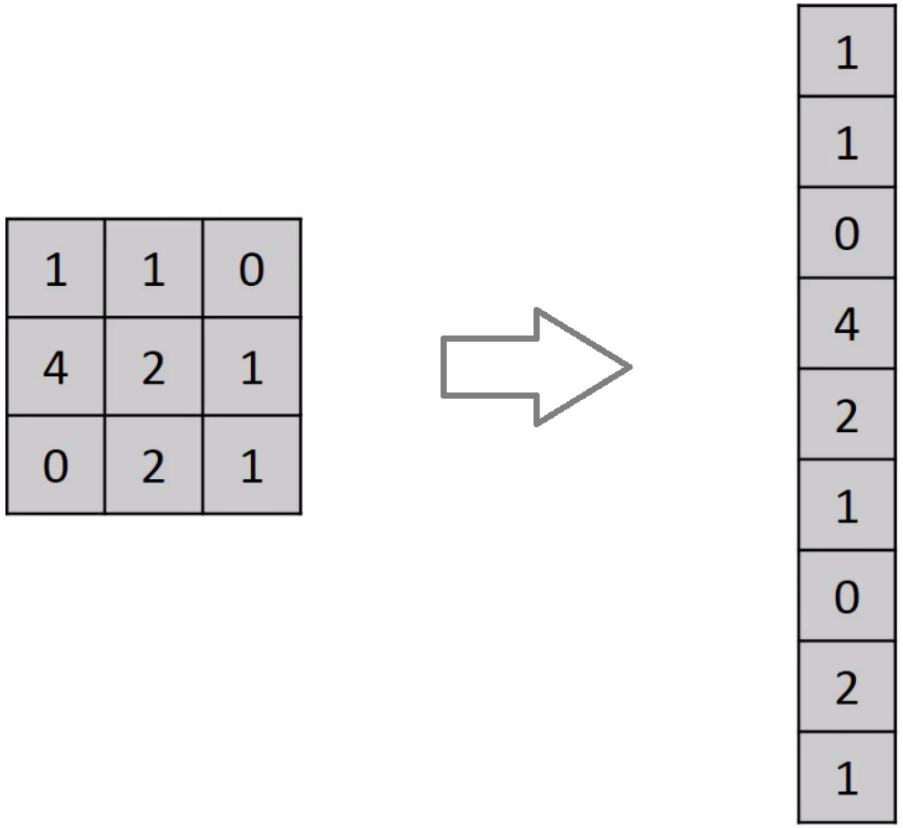
Disable some neurons with probability *p*.
Allows the network to learn different paths for the same solution. Helps avoid overfitting.

Activation Function Layer



The same activation function but as a layer.
We use ReLU for hidden layers and Softmax or Linear for the output layer.

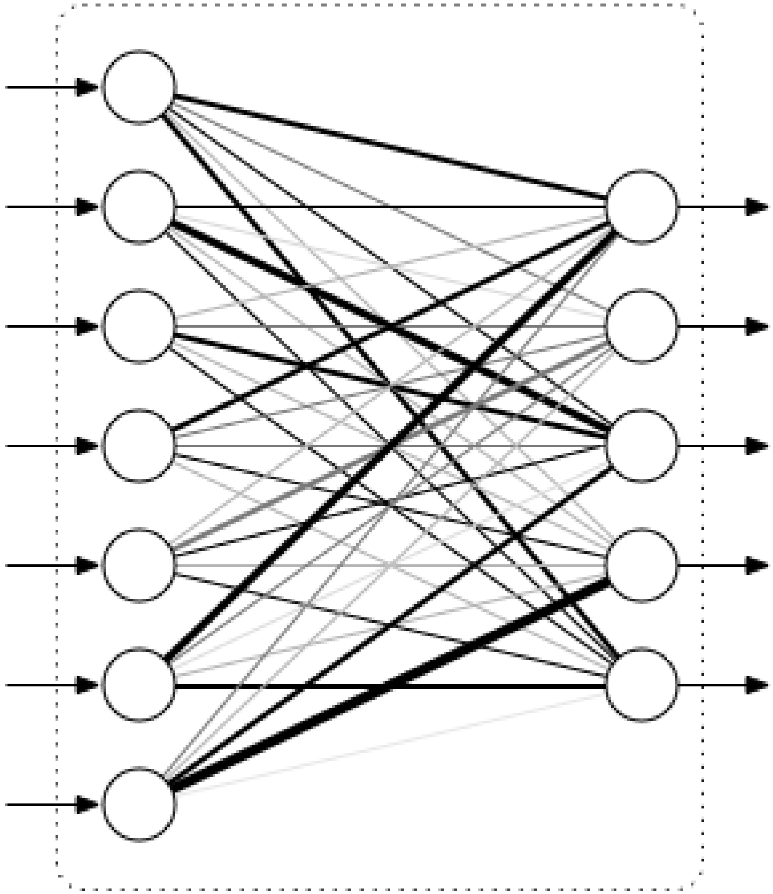
Flatten Layer



$$\begin{bmatrix}1&1&0\\4&2&1\\0&2&1\end{bmatrix} \Rightarrow \begin{bmatrix}1\\1\\0\\4\\2\\1\\0\\2\\1\end{bmatrix}$$

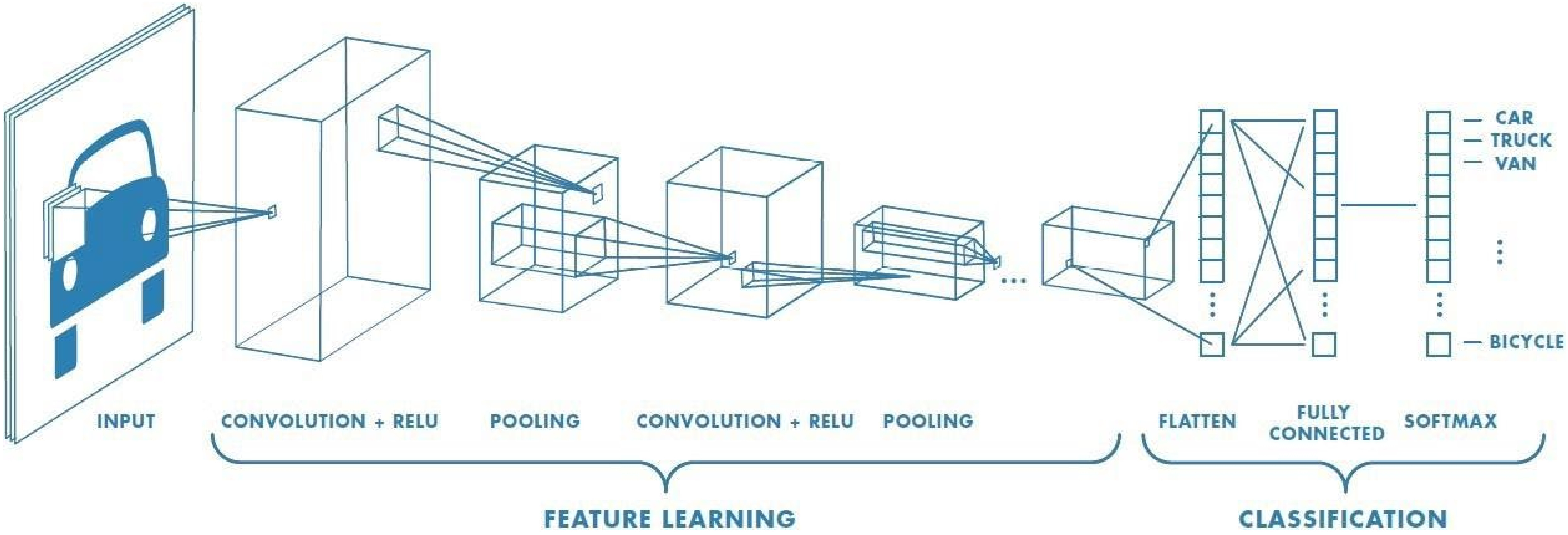
After all feature extractions, we have to make a choice….
The Flatten layer will transform the matrix into an array for the Fully Connected Layer.

Fully Connected Layer



We use a FCN as a layer to make our prediction.

Convolutional Neural Network (CNN)



Best architecture to use when we are dealing with image projects.

End of Lecture 1

Next up Part 3 Lecture 2: Characterization and Monitoring