An introduction to Deep Learning

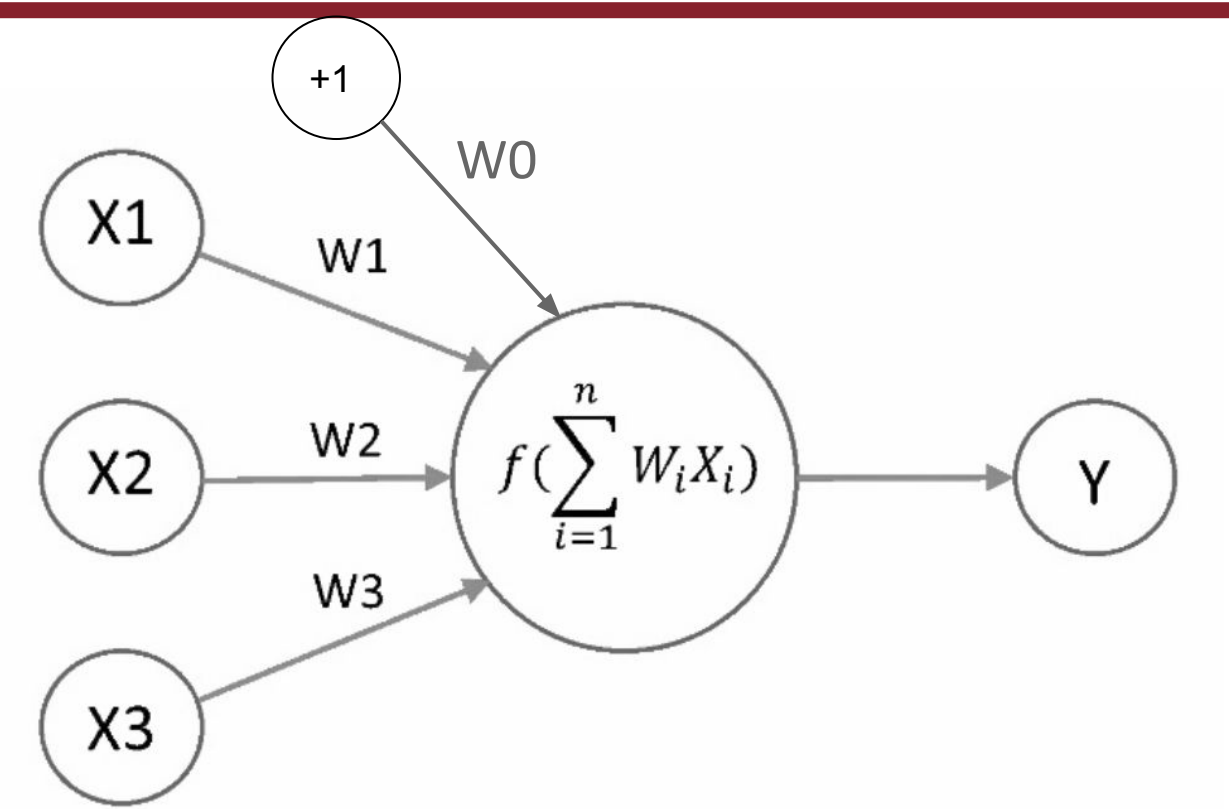
Pedro Lara Benítez

Manuel Carranza García

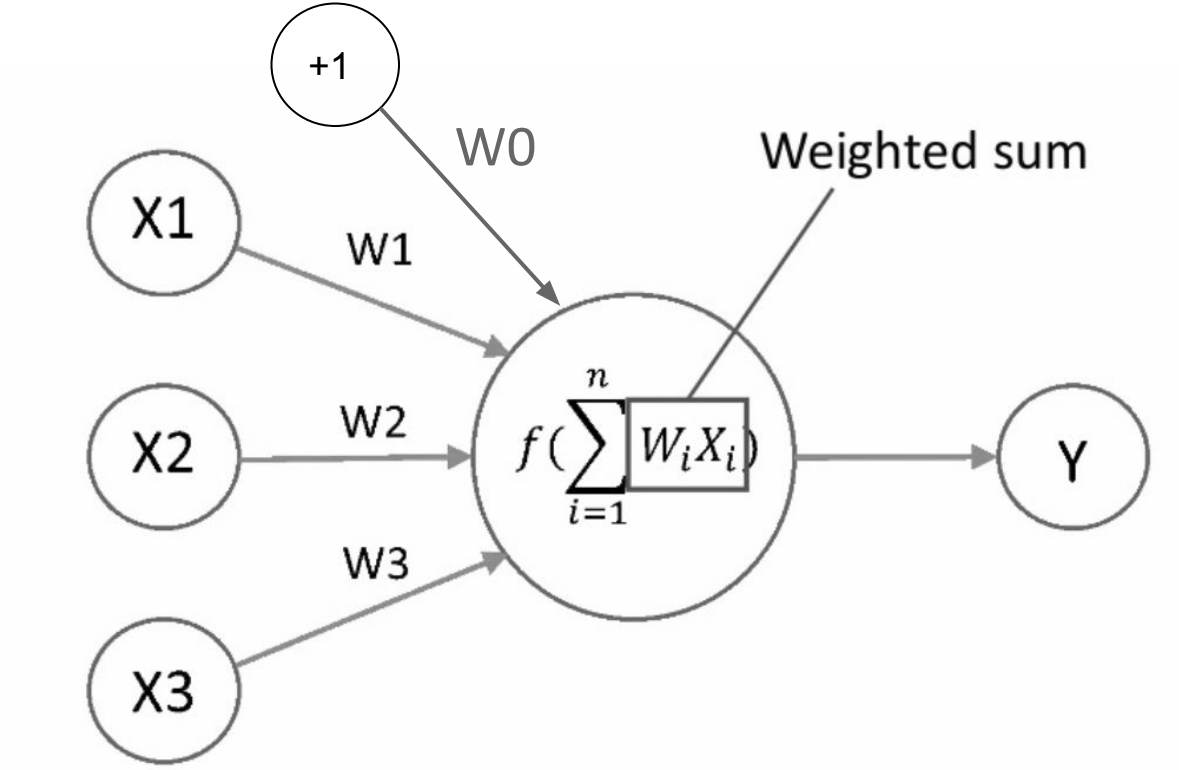
Fundamental concepts

- Neurons
- Multi-layer perceptron
- Training a NN

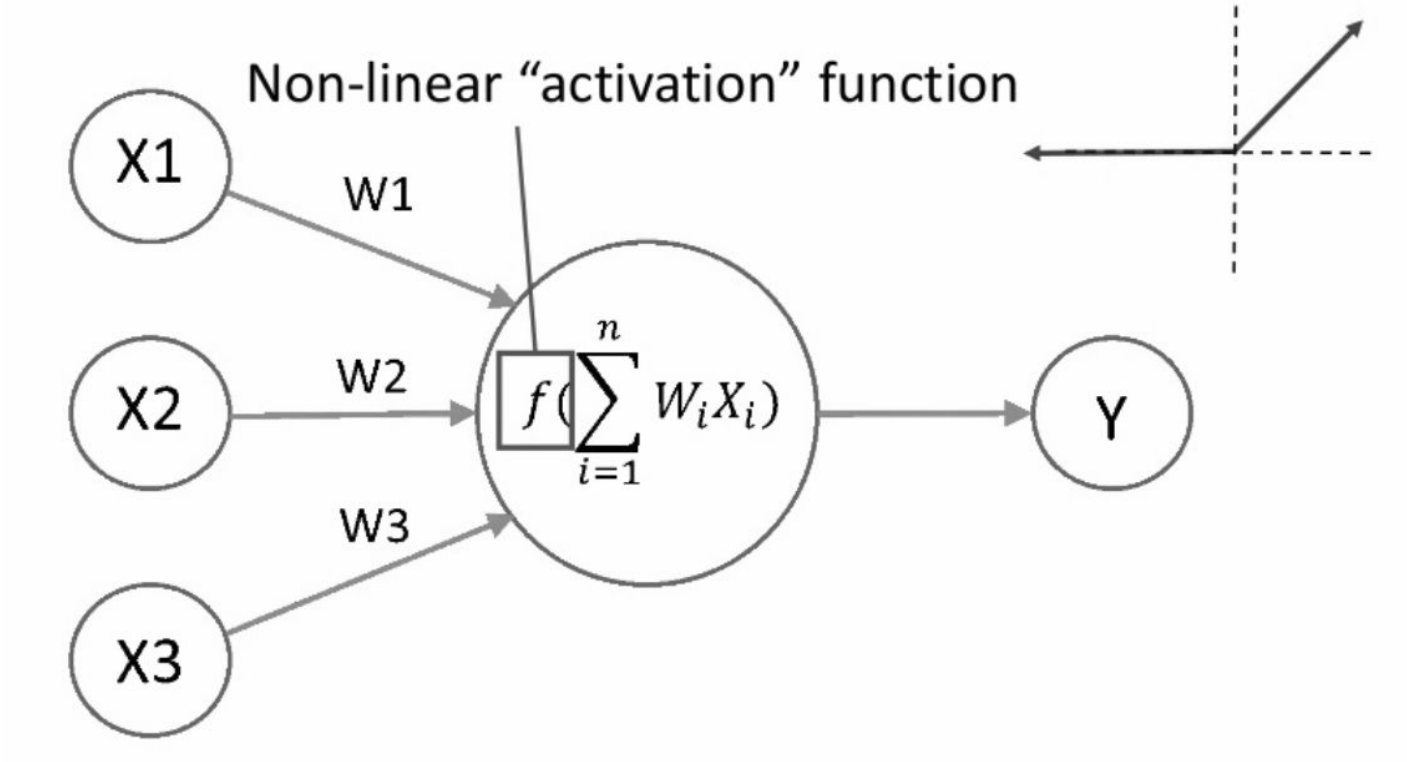
Perceptron



Perceptron



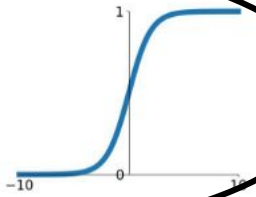
Perceptron


Non-linear “activation” function
X1
X2
X3
W1
W2
W3
$f(\sum_{i=1}^{n} W_i X_i)$
Y

Activation functions

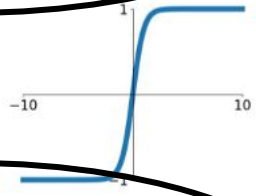
Sigmoid

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



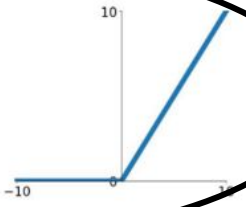
tanh

$$\tanh(x)$$



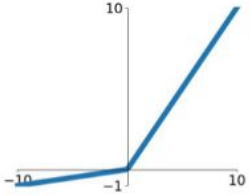
ReLU

$$\max(0, x)$$



Leaky ReLU

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

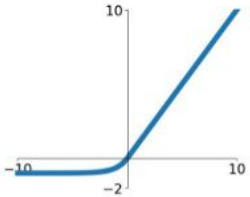


Maxout

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

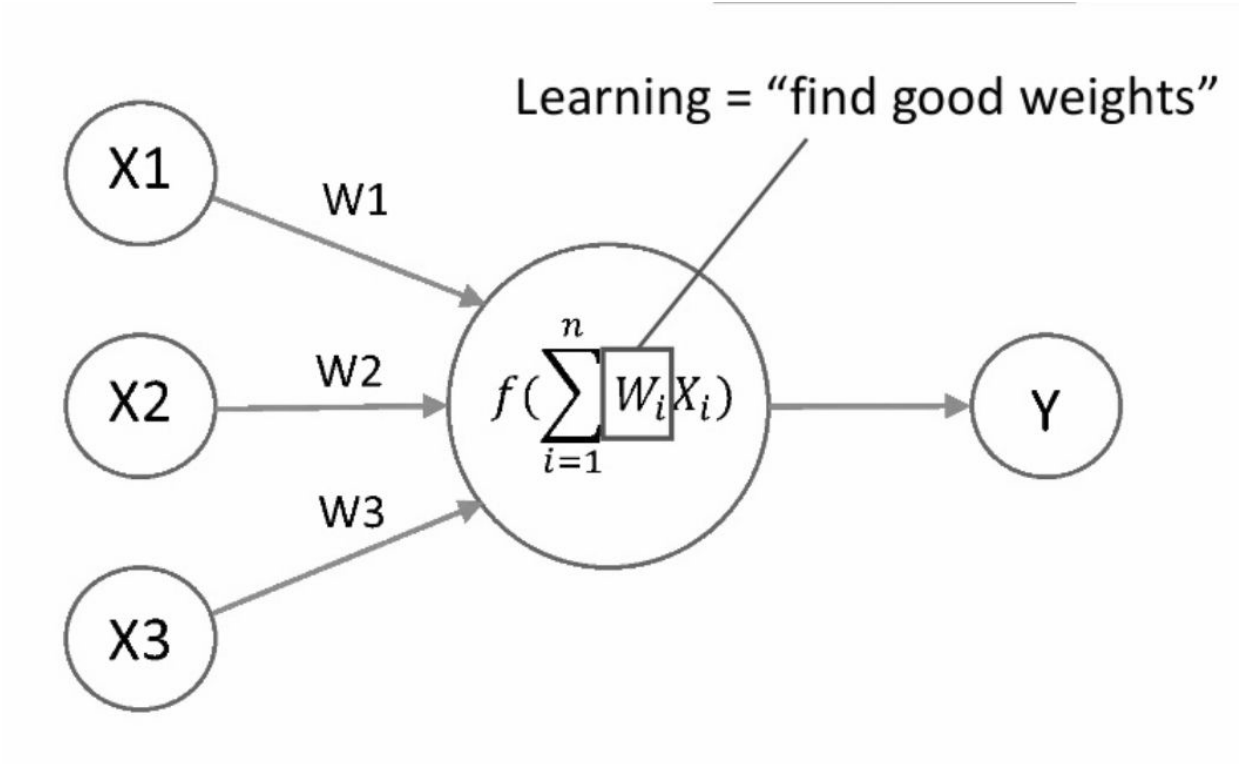
ELU

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

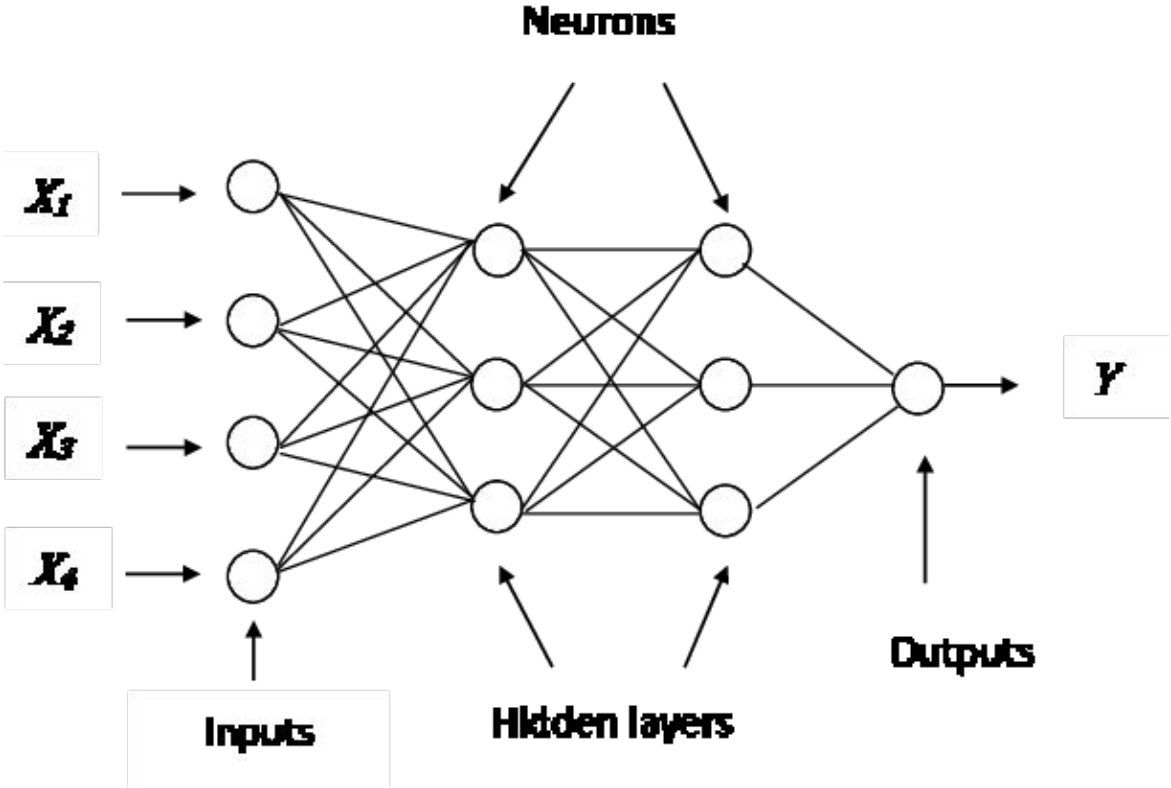


Softmax for multi-class classification

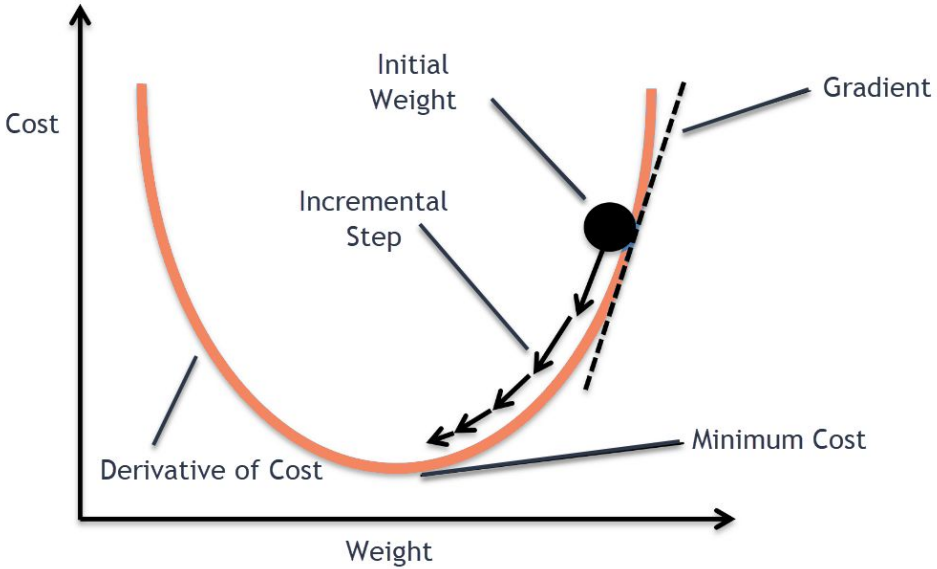
Perceptron



Multi-layer Perceptron



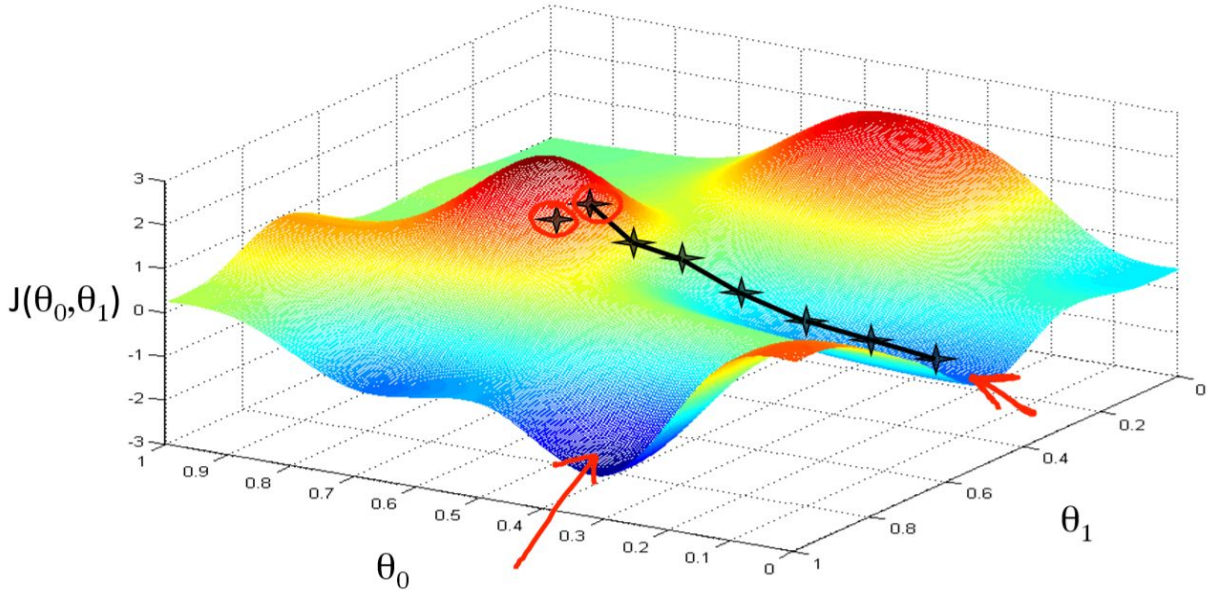
Training a NN



Optimize a cost function using

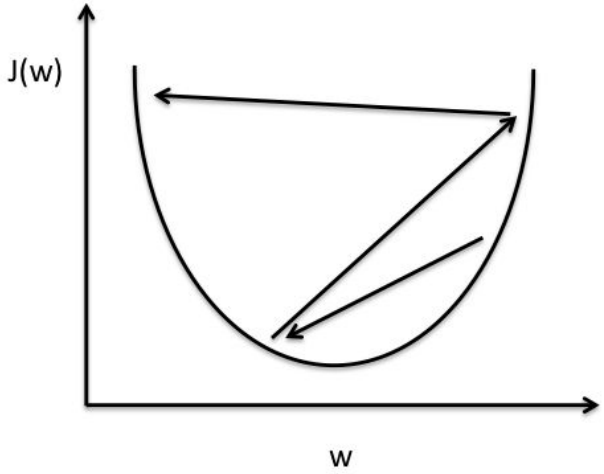
Gradient Descent with Back-Propagation

Training a NN

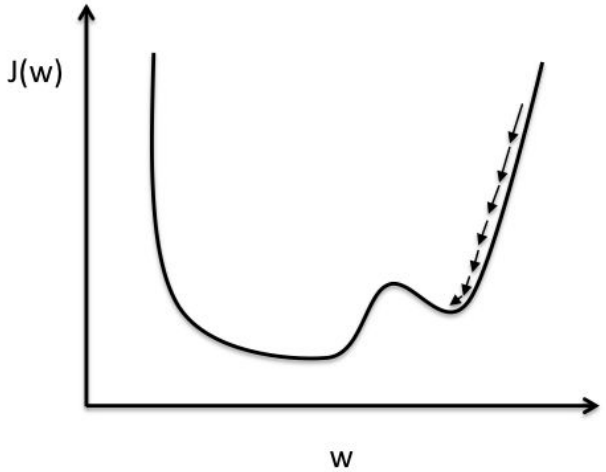


Learning rate

$$w = w + \Delta w, \text{ where } \Delta w = -\eta \nabla J(w)$$

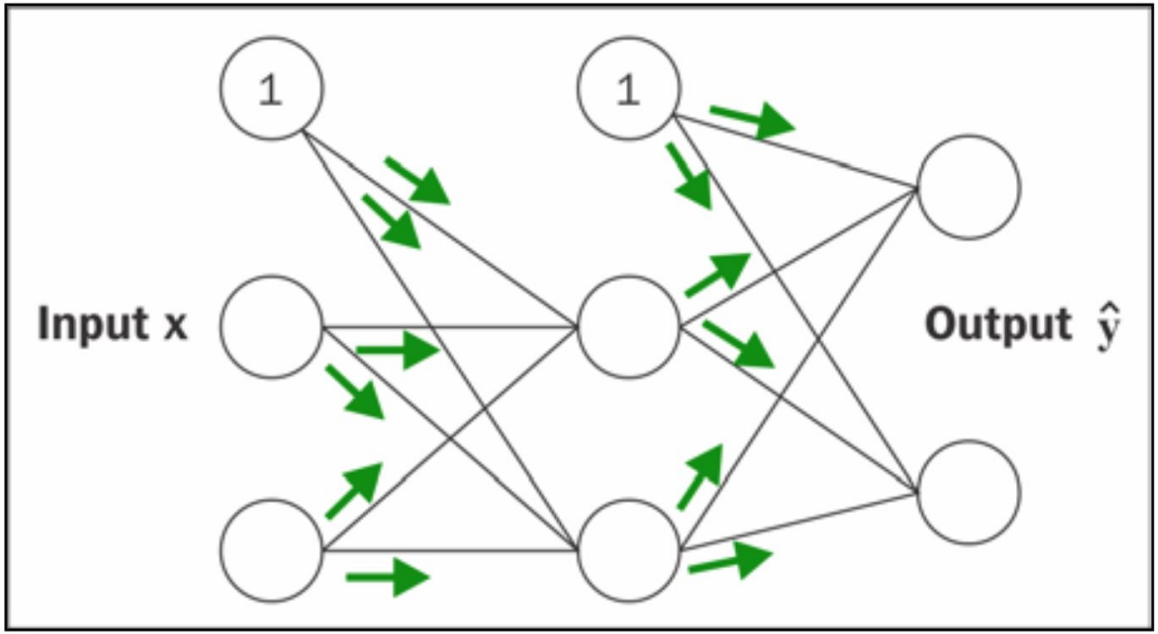


Large learning rate: Overshooting.

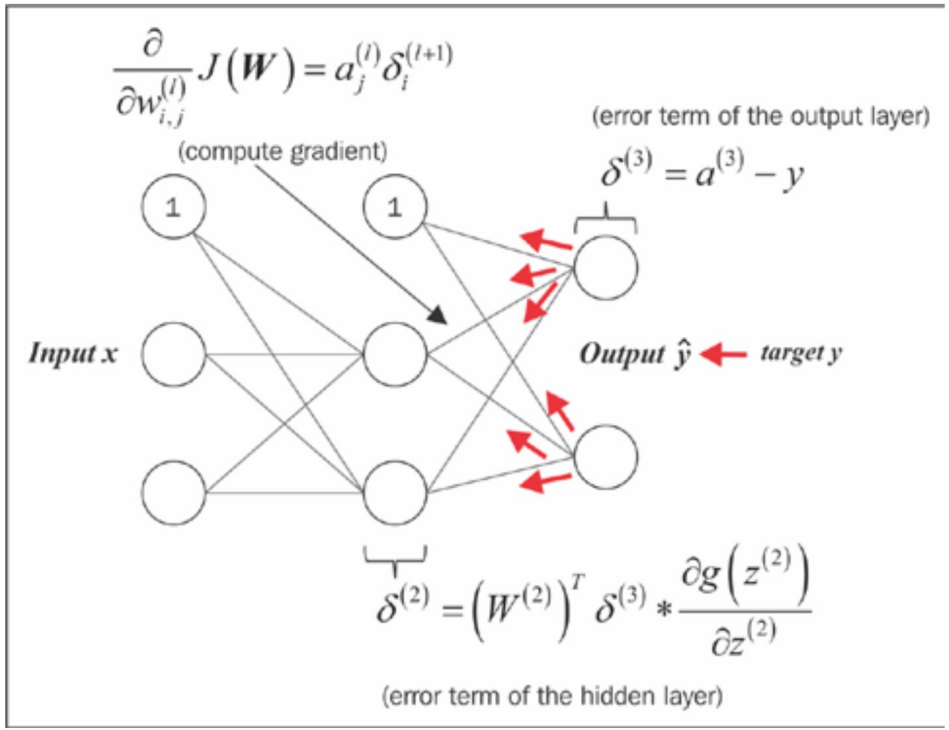


Small learning rate: Many iterations until convergence and trapping in local minima.

Backpropagation - Forward



Backpropagation - Backwards

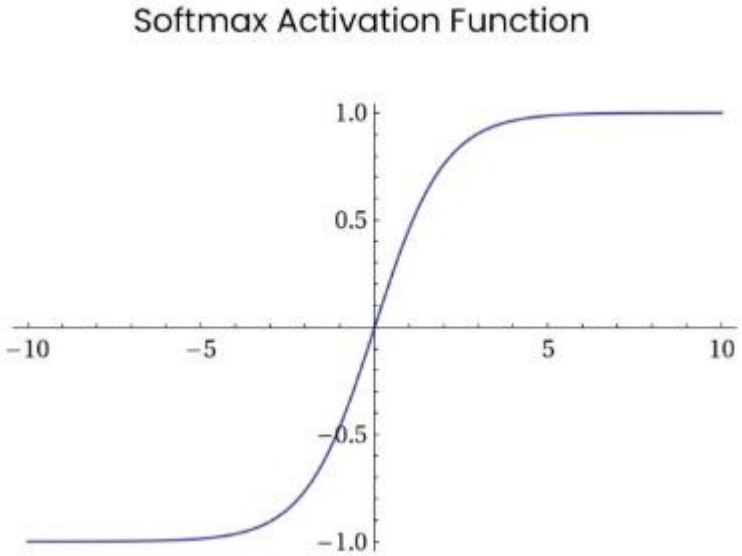


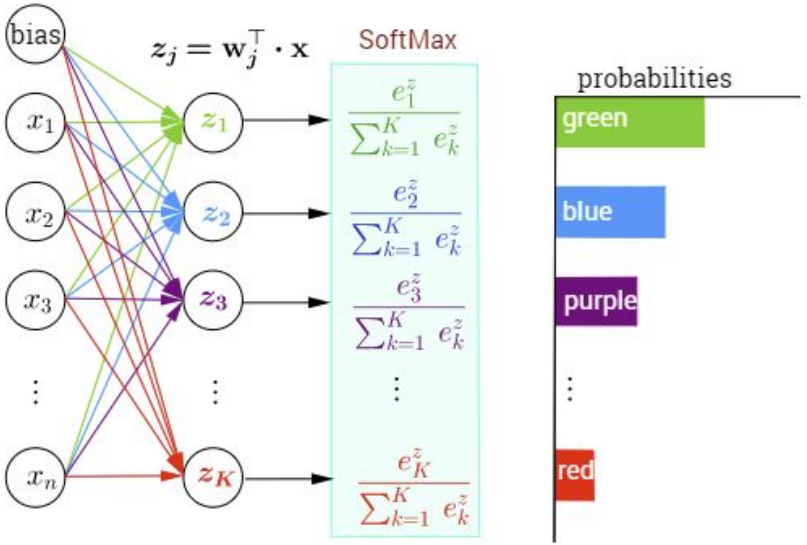
Important terms

- Optimizer: SGD, Adam, RMSProp
- Loss: Mean Squared Error, Cross Entropy
- Learning rate
- Activation function
- Batch size
- Epoch
- Dropout

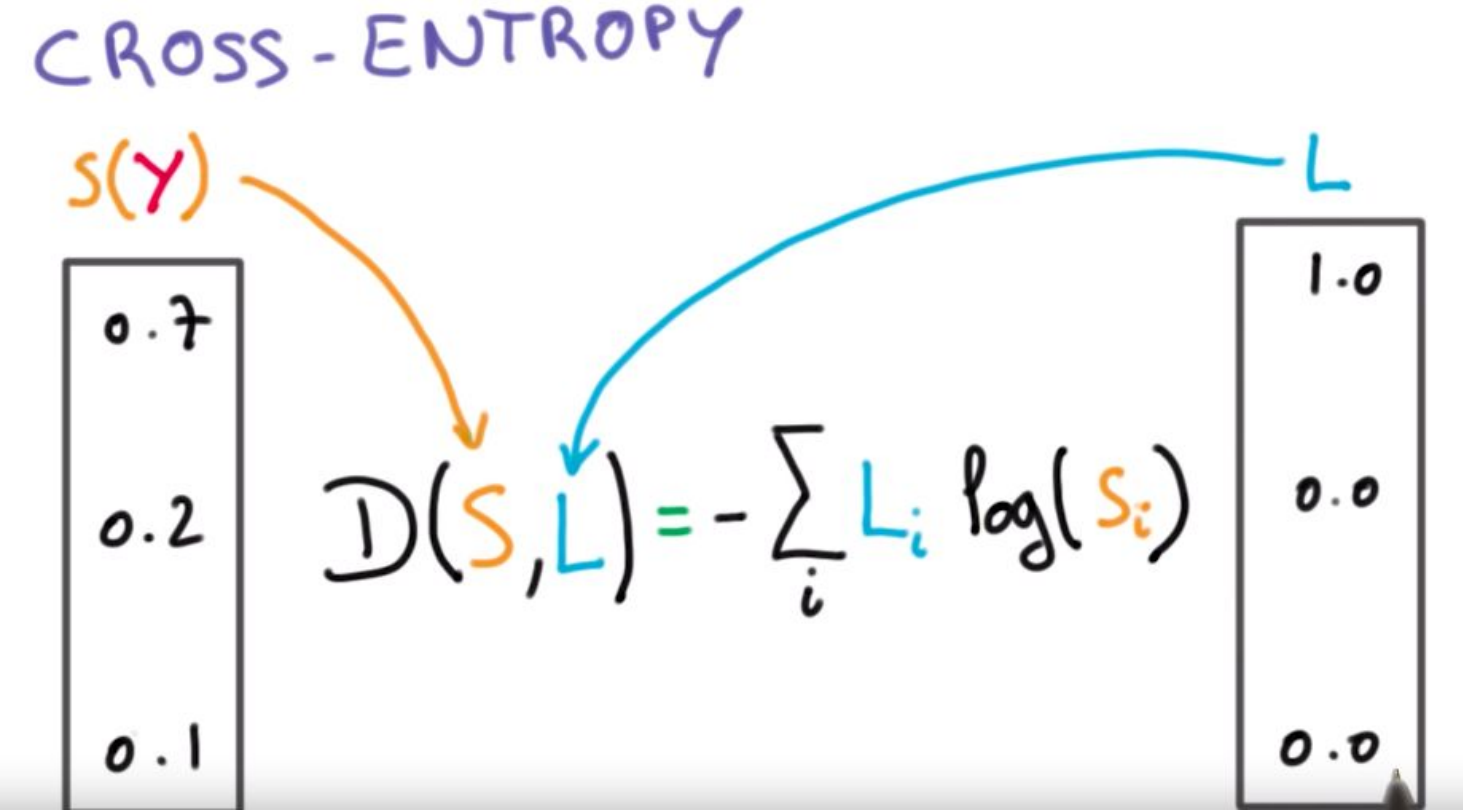
Code example

Softmax activation function

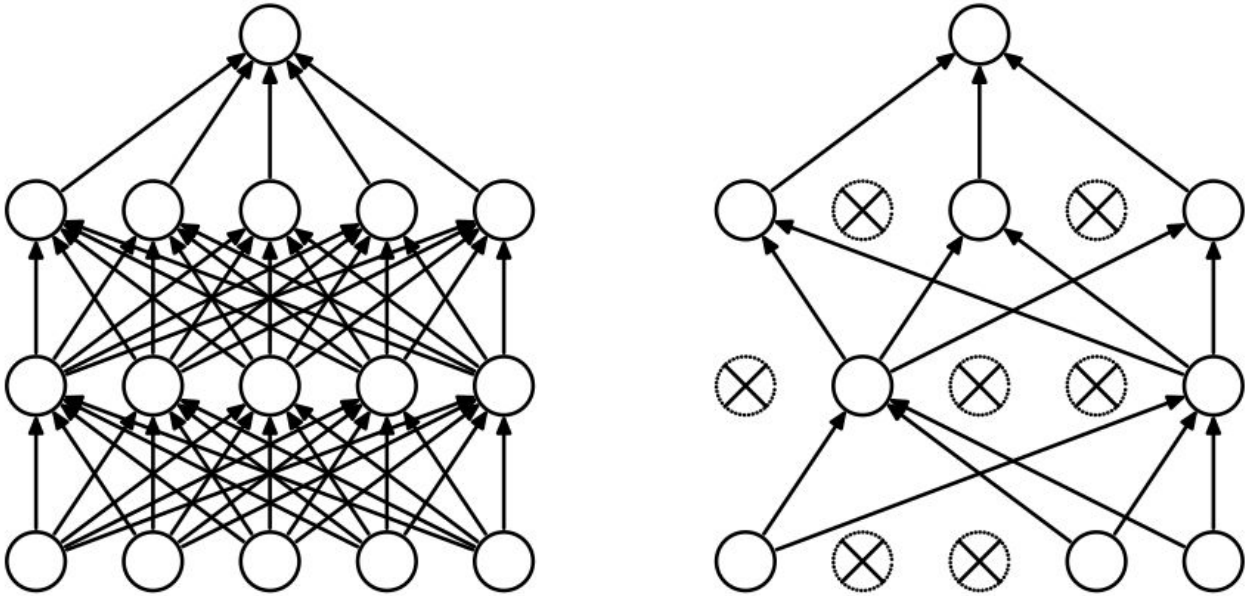




Loss function - cross entropy



Dropout



(a) Standard Neural Net

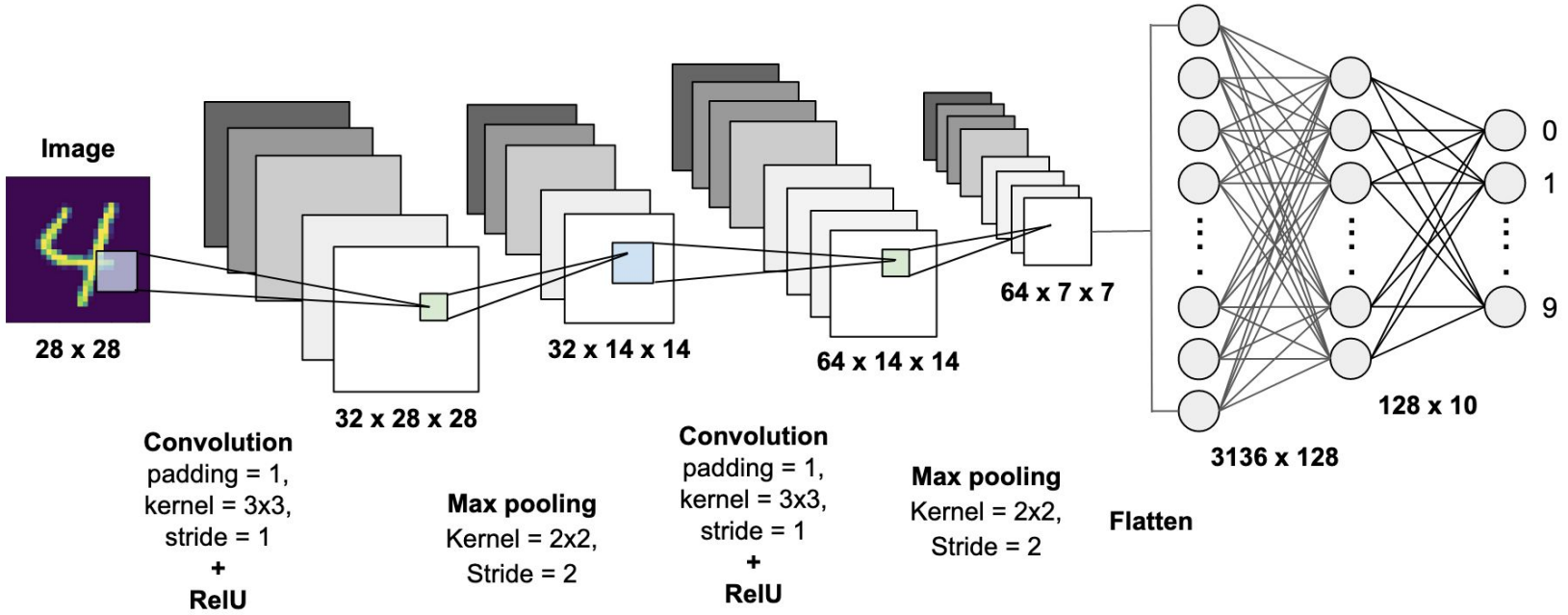
(b) After applying dropout.

Deep learning architectures

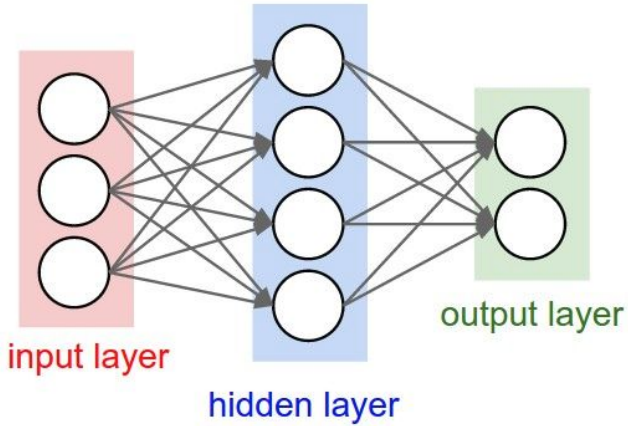
- Convolutional neural networks
- Recurrent neural networks

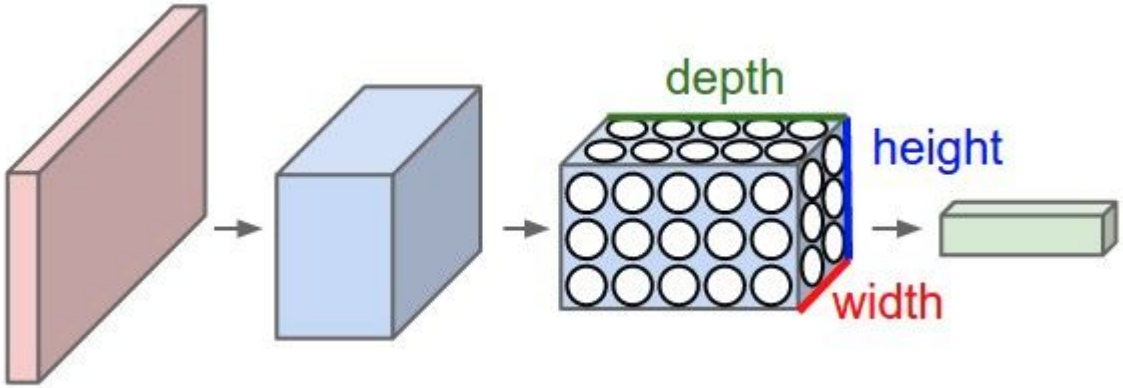
Convolutional Neural Networks

Convolutional NN



MLP vs CNN

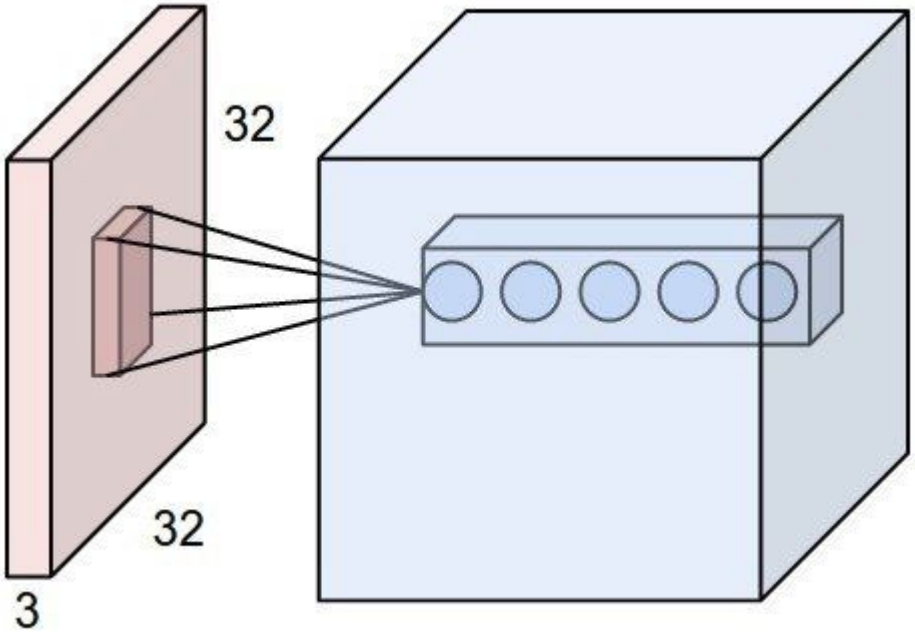




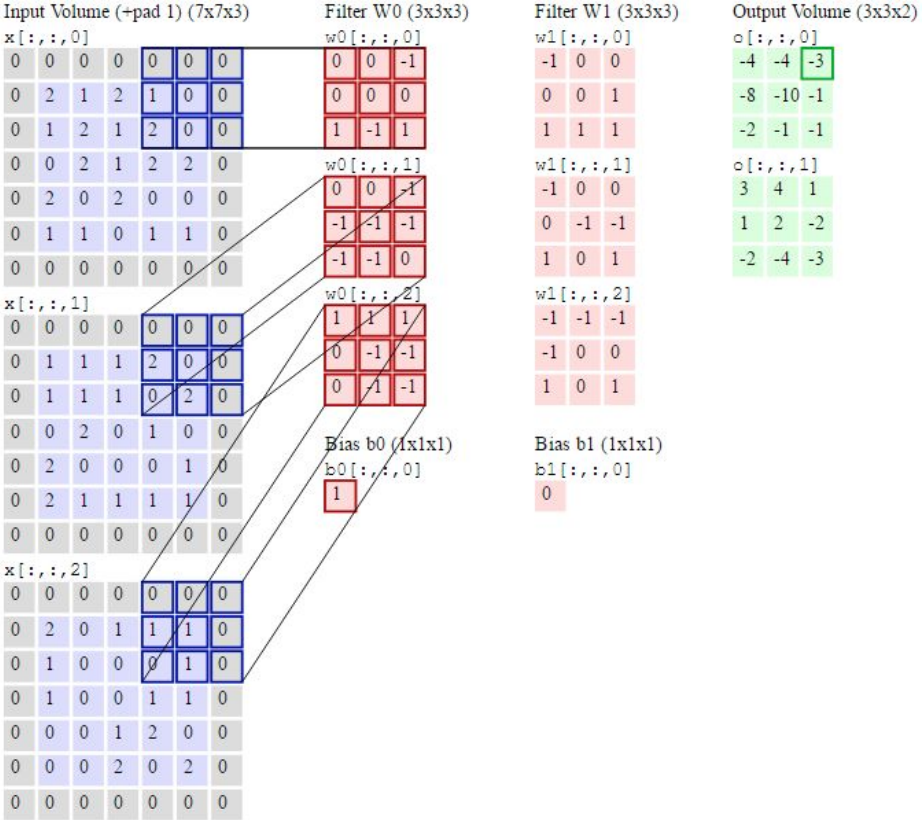
MLP

CNN

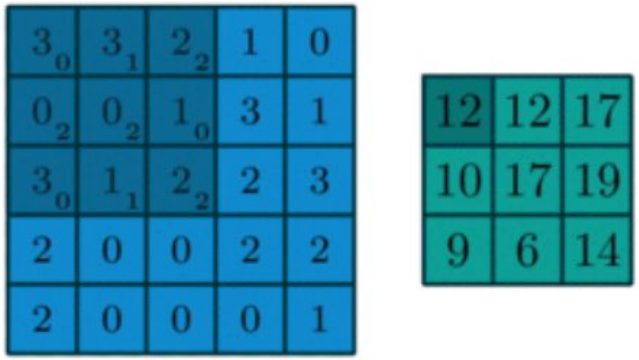
CNN local connectivity



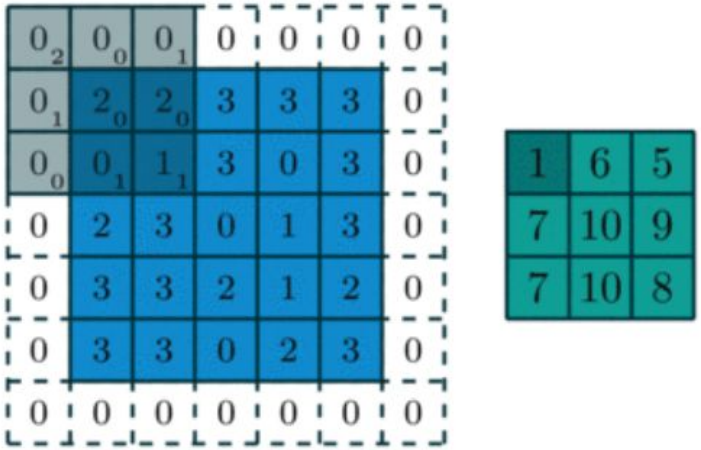
Convolution operation



Padding

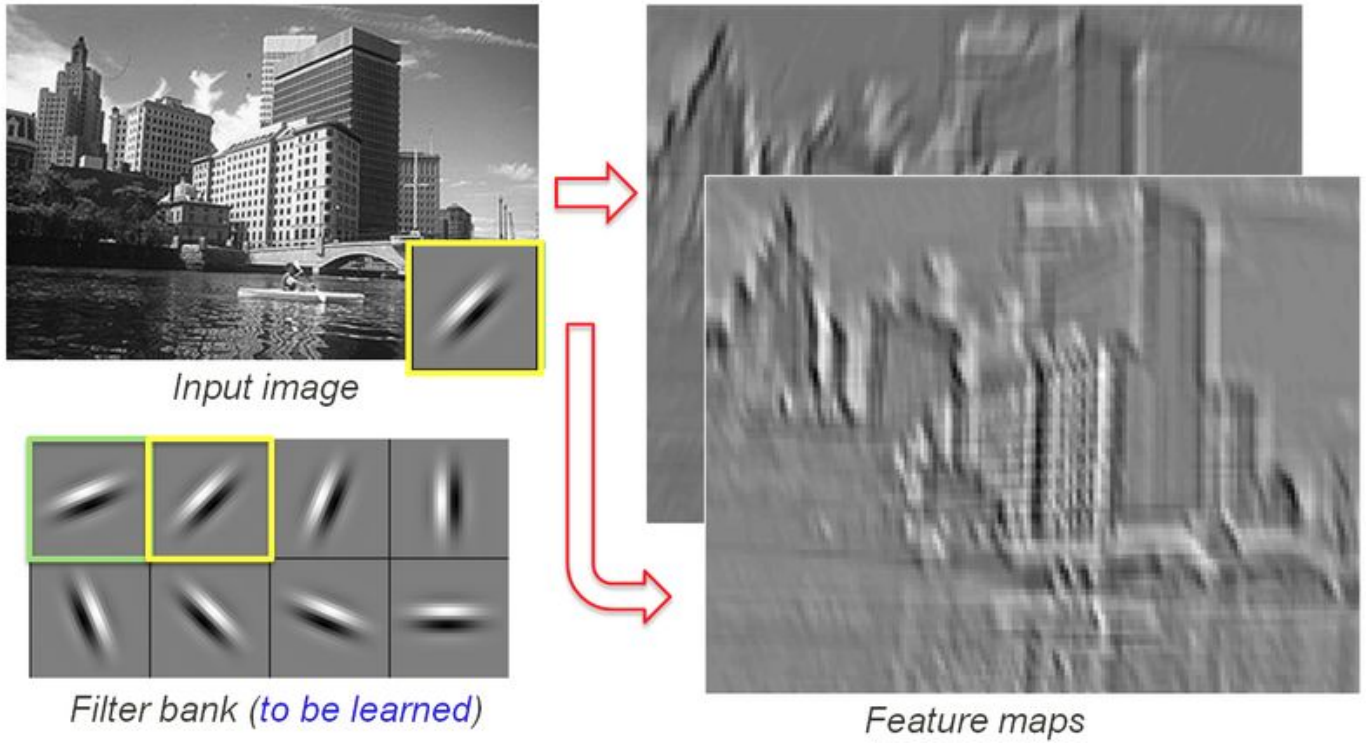


(a) Padding=0, Stride=1

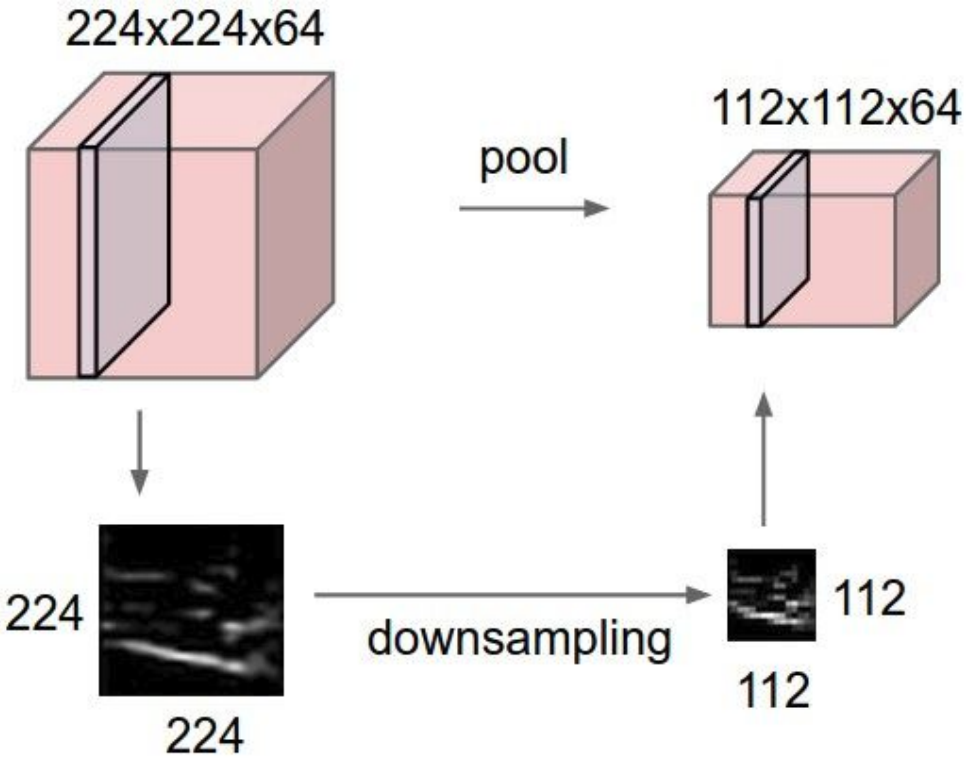


(b) Padding=1, Stride=2

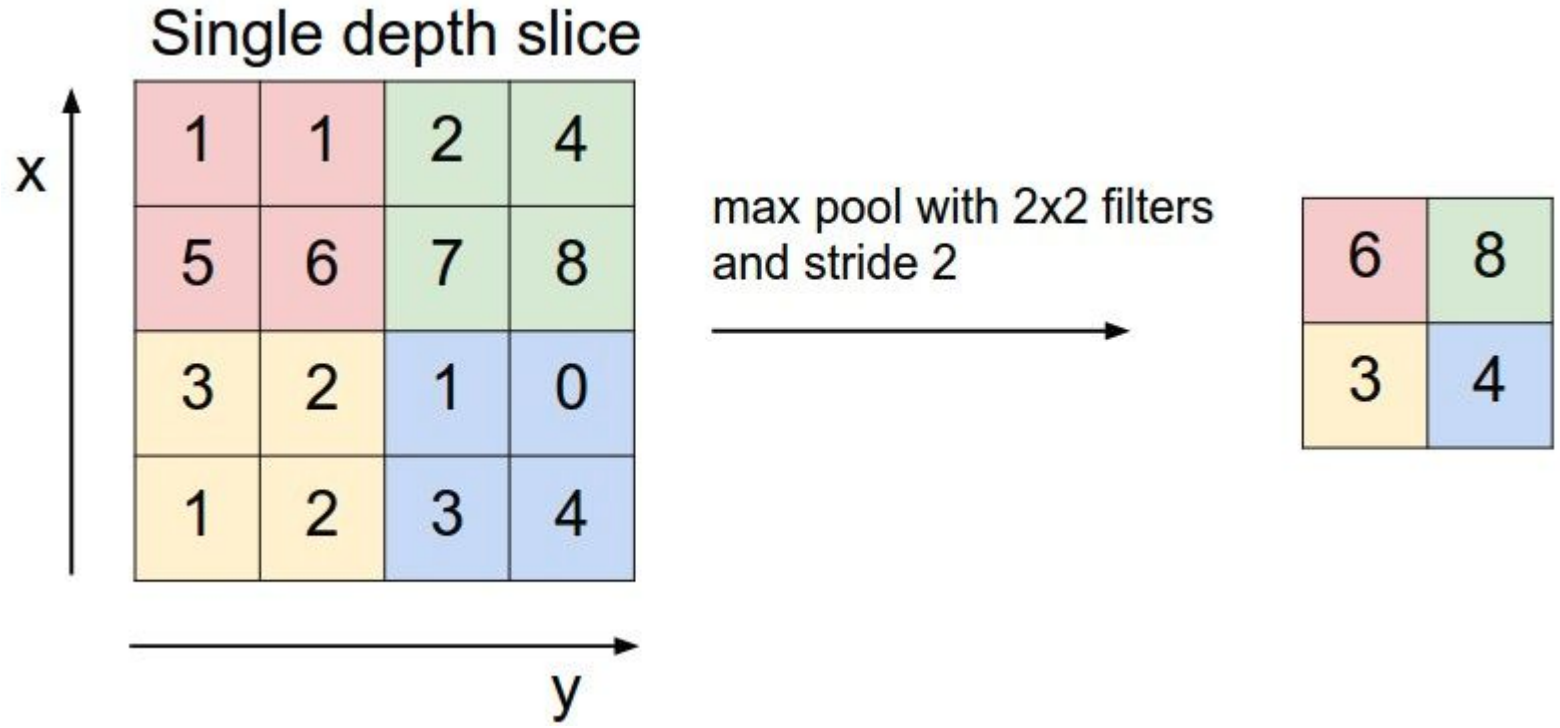
Convolutional feature maps



Pooling



Max-Pooling


Single depth slice
x
y
1 1 2 4
5 6 7 8
3 2 1 0
1 2 3 4
max pool with 2x2 filters
and stride 2
6 8
3 4

Important terms (CNN)

- Feature maps = Filters
- Kernel size
- Stride
- Padding
- Pool size