

Neural Networks

Declan Groves

June 30 2016

Outline

Neural Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

- 1 What is
- 2 Gradient
- 3 Backprop
- 4 Activations
- 5 How to
- 6 Convolution

Overview

Neural
Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

- Hypest ML
- Good at unstructured problems
- Suboptimal at structured problems
- cs231n.github.io
- github.com/DexGroves/nn-intro (references.md)

History

Neural Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

- Around since the 1950s
- Resurgence in 1970s
- Resurgence in late 2000s

A graphical linear model

Neural
Networks

Declan Groves

What is

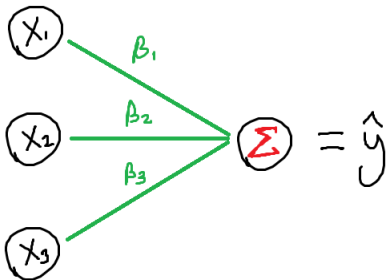
Gradient

Backprop

Activations

How to

Convolution



A graphical linear model

Neural
Networks

Declan Groves

What is

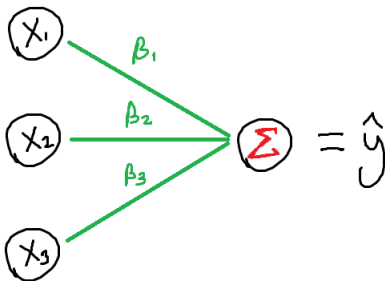
Gradient

Backprop

Activations

How to

Convolution



$$X W = \hat{y}$$

A graphical linear model

Neural
Networks

Declan Groves

What is

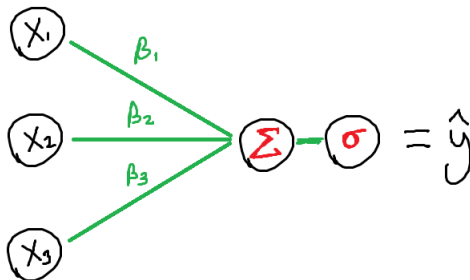
Gradient

Backprop

Activations

How to

Convolution



$$\sigma(XW) = \hat{y}$$

A graphical linear model

Neural
Networks

Declan Groves

What is

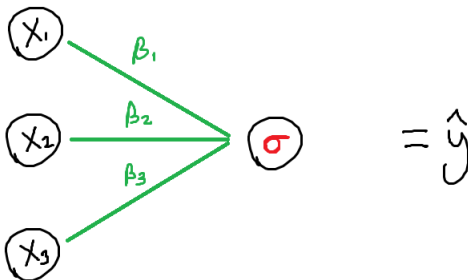
Gradient

Backprop

Activations

How to

Convolution



$$\sigma(XW) = \hat{y}$$

A simple neural network

Neural
Networks

Declan Groves

What is

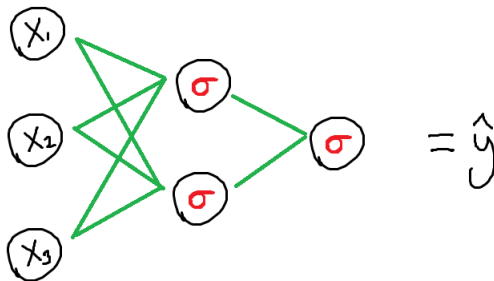
Gradient

Backprop

Activations

How to

Convolution



A simple neural network

Neural
Networks

Declan Groves

What is

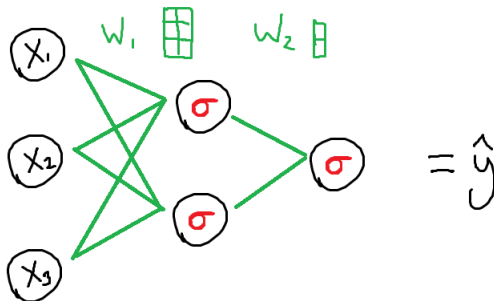
Gradient

Backprop

Activations

How to

Convolution



A simple neural network

Neural
Networks

Declan Groves

What is

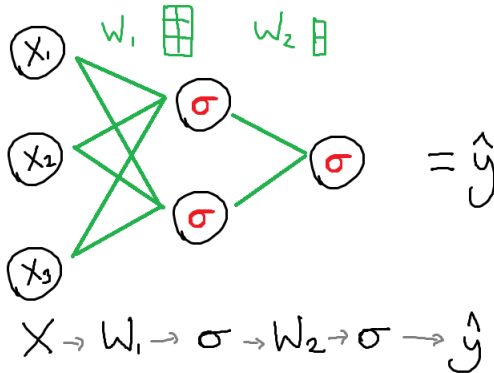
Gradient

Backprop

Activations

How to

Convolution



A simple neural network

Neural Networks

Declan Groves

What is

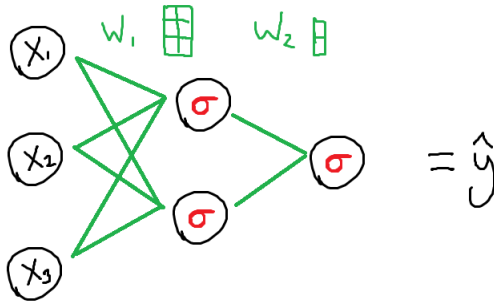
Gradient

Backprop

Activations

How to

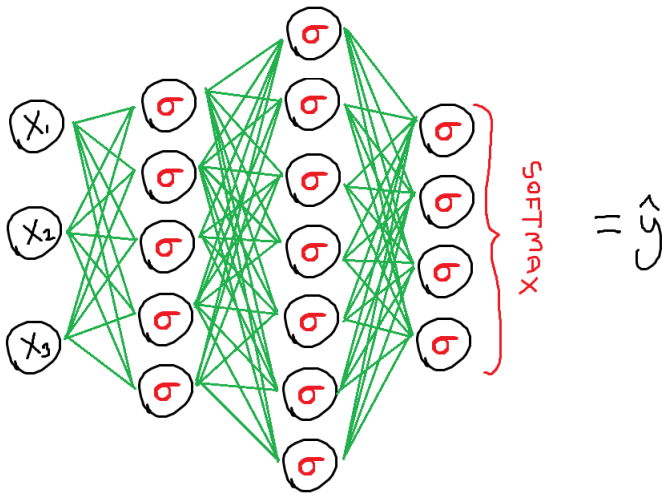
Convolution



$$X \rightarrow W_1 \rightarrow \sigma \rightarrow W_2 \rightarrow \sigma \rightarrow \hat{y}$$

$$\sigma(\sigma(XW_1)W_2) = \hat{y}$$

A multilayer multinomial classifier



Neural
Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

Gradient Descent

Neural
Networks

Declan Groves

What is

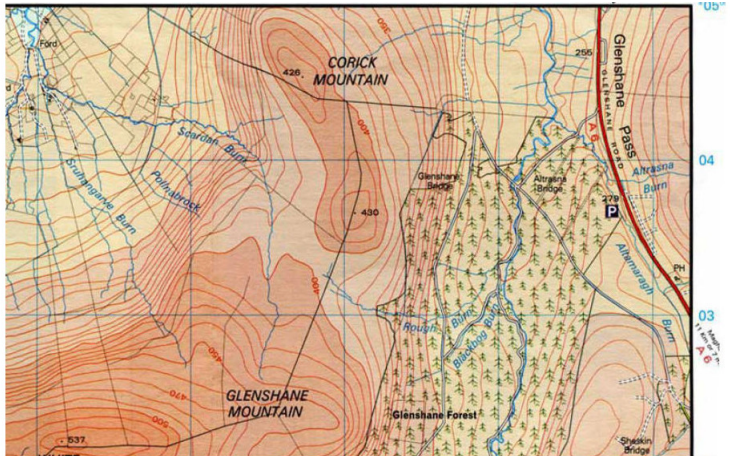
Gradient

Backprop

Activations

How to

Convolution

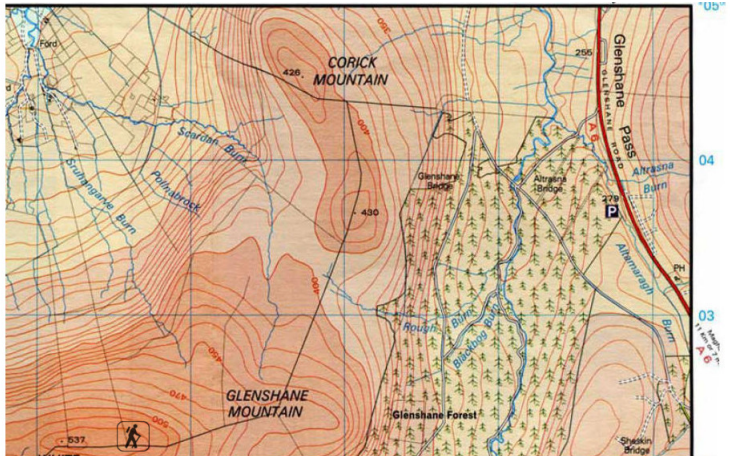


Gradient Descent

Neural Networks

Declan Groves

Gradient



Gradient Descent

Neural
Networks

Declan Groves

What is

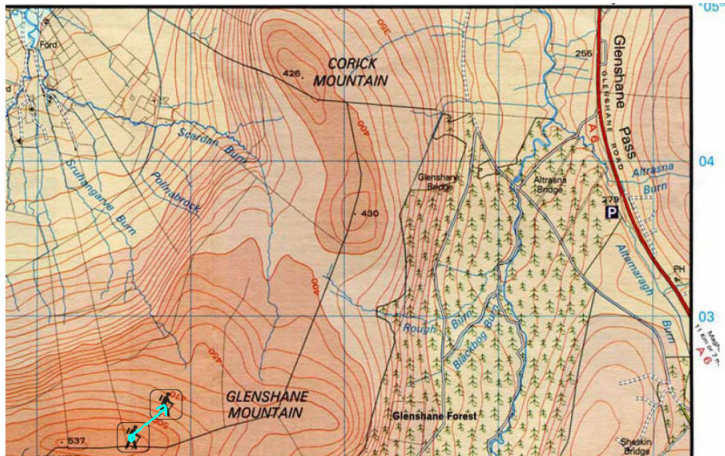
Gradient

Backprop

Activations

How to

Convolution



$$L_{N+1} = L_N - \gamma \nabla f(L_N)$$

Gradient Descent

Neural
Networks

Declan Groves

What is

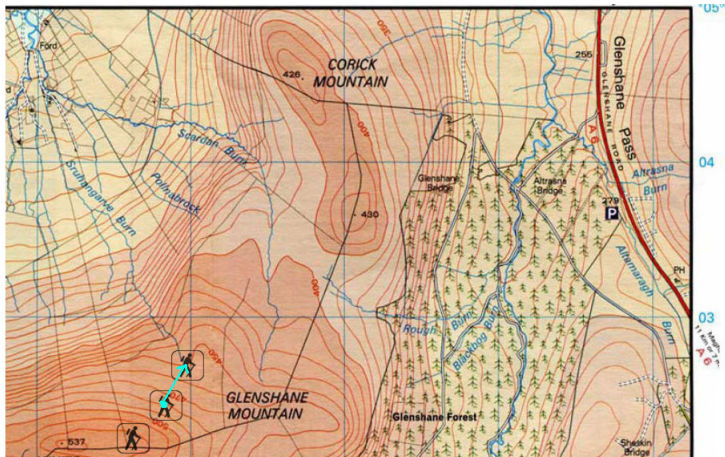
Gradient

Backprop

Activations

How to

Convolution



$$L_{N+1} = L_N - \gamma \nabla f(L_N)$$

Gradient Descent

Neural
Networks

Declan Groves

What is

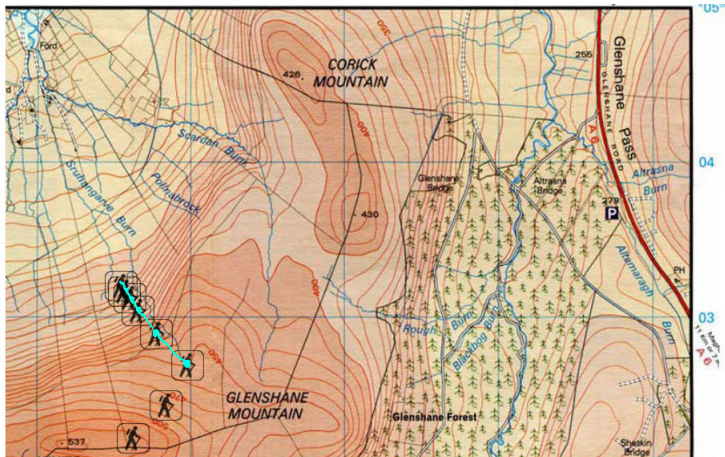
Gradient

Backprop

Activations

How to

Convolution



$$L_{N+1} = L_N - \gamma \nabla f(L_N)$$

Gradient Descent and Momentum

Neural
Networks

Declan Groves

What is

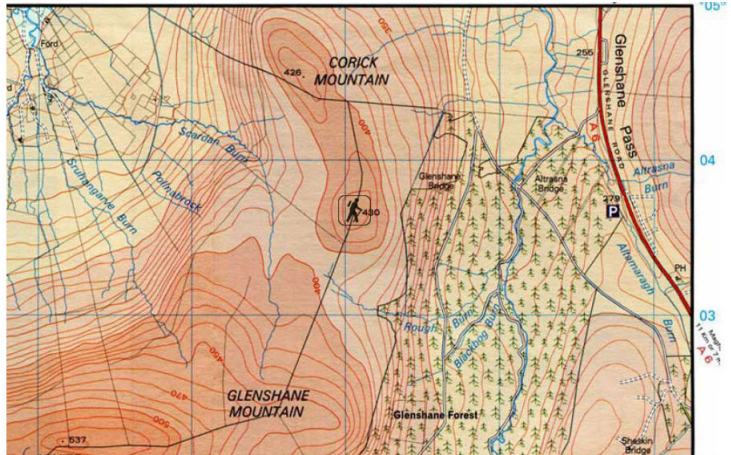
Gradient

Backprop

Activations

How to

Convolution



Gradient Descent and Momentum

Neural
Networks

Declan Groves

What is

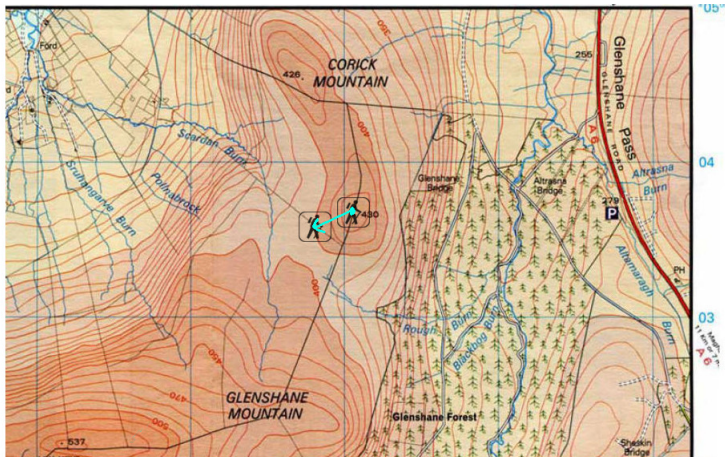
Gradient

Backprop

Activations

How to

Convolution



Gradient Descent and Momentum

Neural
Networks

Declan Groves

What is

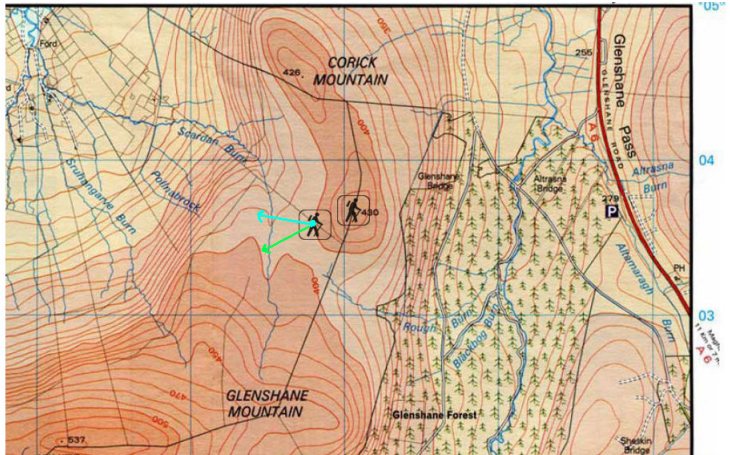
Gradient

Backprop

Activations

How to

Convolution



$$L_{N+1} = L_N - \gamma \nabla f(L_N) - m \gamma \nabla f(L_{N-1})$$

Gradient Descent and Momentum

Neural
Networks

Declan Groves

What is

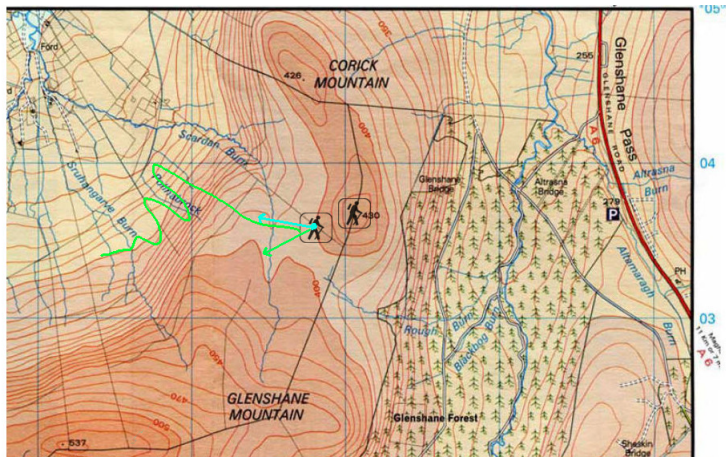
Gradient

Backprop

Activations

How to

Convolution



$$L_{N+1} = L_N - \gamma \nabla f(L_N) - m\gamma \nabla f(L_{N-1})$$

Ravines

Neural Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution



Ravines - without momentum

Neural
Networks

Declan Groves

What is

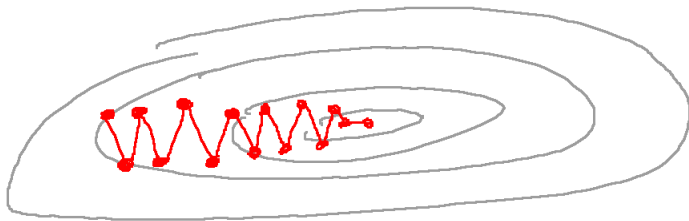
Gradient

Backprop

Activations

How to

Convolution



Ravines - with momentum

Neural
Networks

Declan Groves

What is

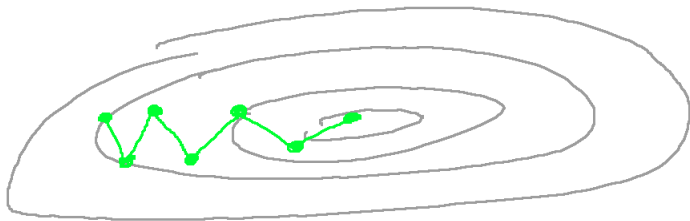
Gradient

Backprop

Activations

How to

Convolution



Nesterov momentum

Neural
Networks

Declan Groves

What is

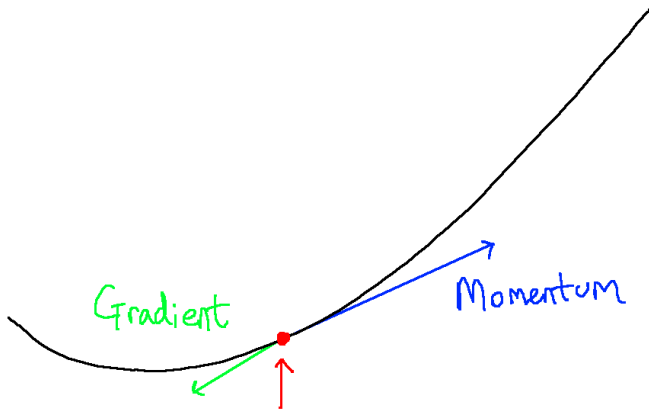
Gradient

Backprop

Activations

How to

Convolution



Nesterov momentum

Neural
Networks

Declan Groves

What is

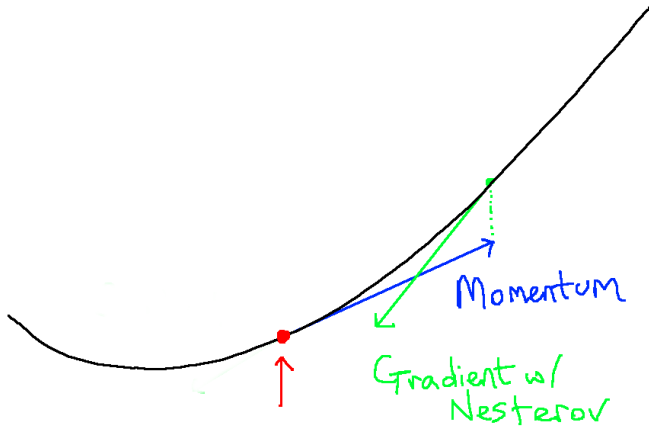
Gradient

Backprop

Activations

How to

Convolution



Stochastic gradient descent

Neural
Networks

Declan Groves

What is

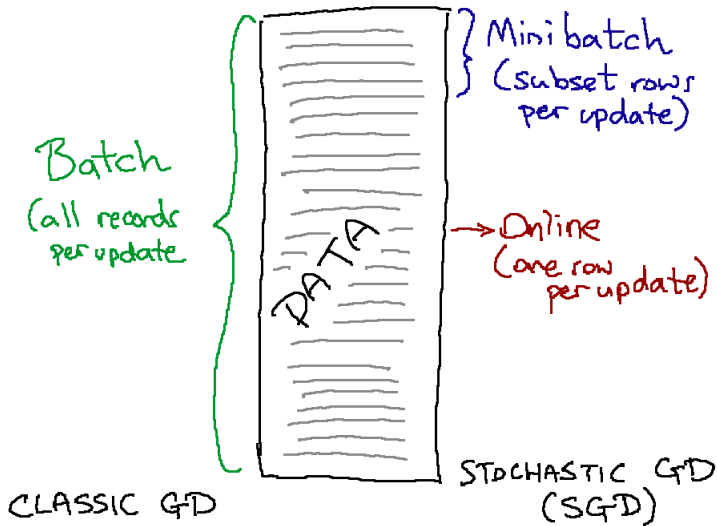
Gradient

Backprop

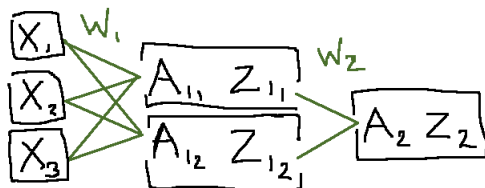
Activations

How to

Convolution



Forward pass



Neural
Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

Forward pass

Neural
Networks

Declan Groves

What is

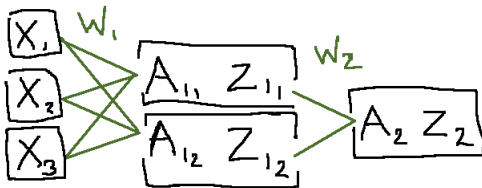
Gradient

Backprop

Activations

How to

Convolution



$$\left. \begin{array}{l} A_1 = XW_1 \\ Z_1 = \sigma(A_1) \\ A_2 = Z_1W_2 \\ Z_2 = \sigma(A_2) = \hat{y} \end{array} \right\} \text{Forward pass}$$

Backwards pass

Neural
Networks

Declan Groves

What is

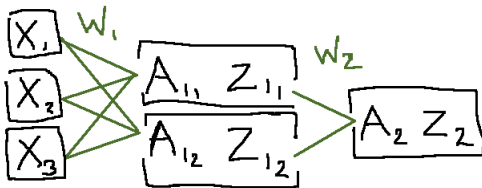
Gradient

Backprop

Activations

How to

Convolution



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

Backwards pass

Neural
Networks

Declan Groves

What is

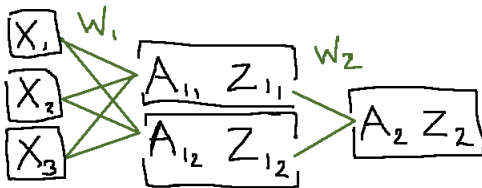
Gradient

Backprop

Activations

How to

Convolution



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\partial J / \partial w_2 =$$

Backwards pass

Neural
Networks

Declan Groves

What is

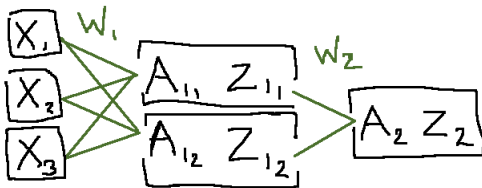
Gradient

Backprop

Activations

How to

Convolution



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\partial J / \partial w_2 = \partial J / \partial \hat{y} \quad \partial \hat{y} / \partial w_2$$

Backwards pass

Neural
Networks

Declan Groves

What is

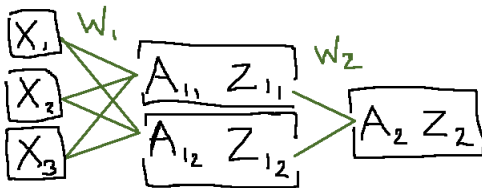
Gradient

Backprop

Activations

How to

Convolution



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\partial J / \partial w_2 = \partial J / \partial \hat{y} \quad \partial \hat{y} / \partial A_2 \quad \partial A_2 / \partial w_2$$

Backwards pass

Neural
Networks

Declan Groves

What is

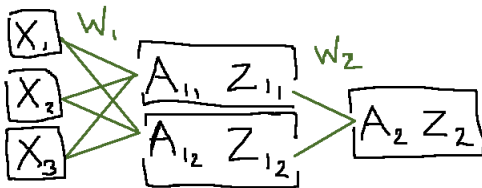
Gradient

Backprop

Activations

How to

Convolution



$$J = \text{err}(y, \hat{y}); J \leftarrow \frac{1}{2}(y - \hat{y})^2$$

$$\begin{aligned} \frac{\partial J}{\partial w_2} &= \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial A_2} \frac{\partial A_2}{\partial w_2} \\ &= \underbrace{(y - \hat{y}) \sigma'(A_2)}_{\delta} Z_1 \end{aligned}$$

Backwards pass

Neural
Networks

Declan Groves

What is

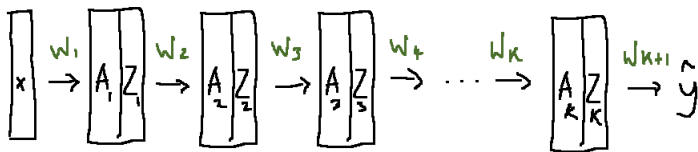
Gradient

Backprop

Activations

How to

Convolution



$$\partial J / \partial w_{i_i} =$$

$$\left(\partial J / \partial \hat{y} \right) \sigma'(A_k) w_k \sigma'(A_{k-1}) w_{k-1} \dots \\ \dots w_{i+1} \sigma'(A_i) Z_{i-1}$$

Backwards pass

Neural
Networks

Declan Groves

What is

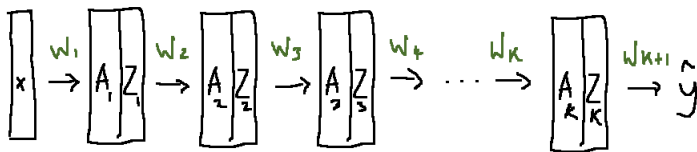
Gradient

Backprop

Activations

How to

Convolution



$$\partial J / \partial w_{i+1} =$$

$$\left(\partial J / \partial \hat{y} \right) \sigma'(A_k) w_k \sigma'(A_{k-1}) w_{k-1} \dots$$

$$\dots w_{i+1} \sigma'(A_i) Z_{i-1} \quad !!$$

Backwards pass

Neural
Networks

Declan Groves

What is

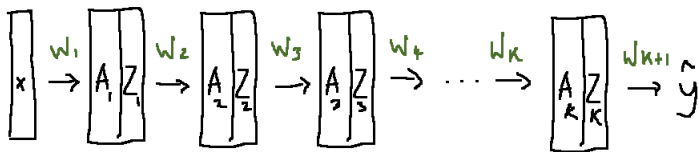
Gradient

Backprop

Activations

How to

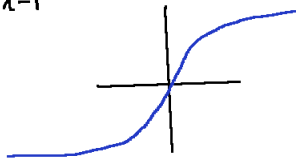
Convolution



$$\partial J / \partial w_{i+1} =$$

$$\left(\partial J / \partial \hat{y} \right) \sigma'(A_k) w_k \sigma'(A_{k-1}) w_{k-1} \dots$$

$$\dots w_{i+1} \sigma'(A_i) Z_{i-1}$$



Linear

Neural
Networks

Declan Groves

What is

Gradient

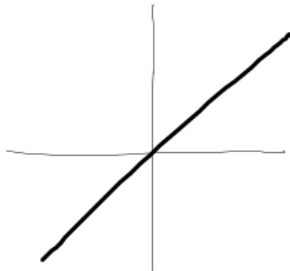
Backprop

Activations

How to

Convolution

■ Useless



Threshold

Neural Networks

Declan Groves

What is

Gradient

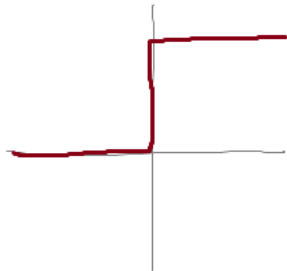
Backprop

Activations

How to

Convolution

- Similar to biological neuron
- No gradient



Sigmoid

Neural
Networks

Declan Groves

What is

Gradient

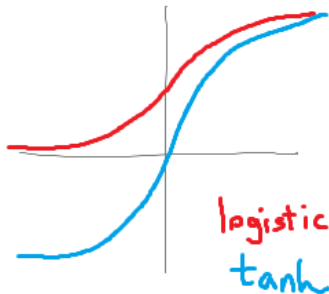
Backprop

Activations

How to

Convolution

- *tanh* preferred
- Gradients can vanish



ReLU

Neural Networks

Declan Groves

What is

Gradient

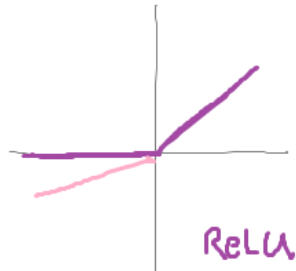
Backprop

Activations

How to

Convolution

- No vanishing gradient
- Cheap to compute
- Can explode and die
- Popular with CNNs



ReLU

Leaky ReLU

Radial basis functions

Neural
Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

- Gaussian + others
- Train very quickly
- Good at interpolation



Six simple rules for training your neural network

Neural
Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

- Use ReLU
- Use minibatch GD with Nesterov momentum (≈ 0.9)
- ≤ 3 layers (unless convolutional)
- Use dropout (≈ 0.2 input layer, ≈ 0.5 else)
- Prefer wider with L2 over smaller
- Be careful with weight initialization!

Libraries for newbs

Neural
Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

- mxnet
- Keras

Convolutional Neural Networks

Neural
Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

- Image recognition killer

Conceptual structure

Neural
Networks

Declan Groves

What is

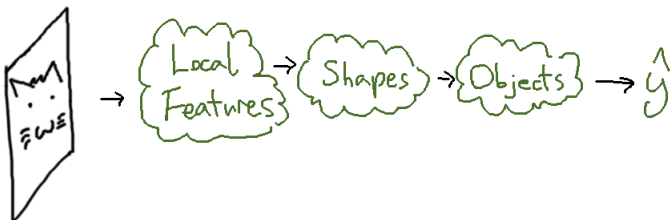
Gradient

Backprop

Activations

How to

Convolution



MLP \rightarrow too many weights!

Neural
Networks

Declan Groves

What is

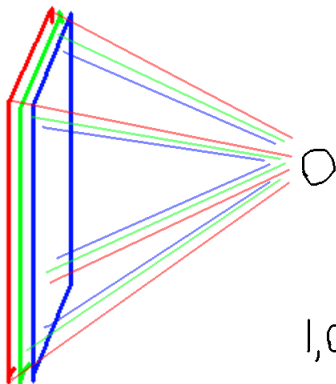
Gradient

Backprop

Activations

How to

Convolution



$$= 3,072 \text{ weights } N^{-1}$$

1,024 Neurons
 \rightarrow $> 3M$ weights

Local connectivity

Neural Networks

Declan Groves

What is

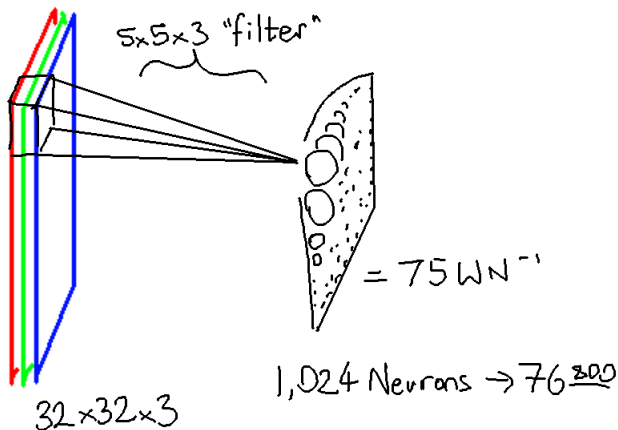
Gradient

Backprop

Activations

How to

Convolution



Local connectivity + convolution

Neural
Networks

Declan Groves

What is

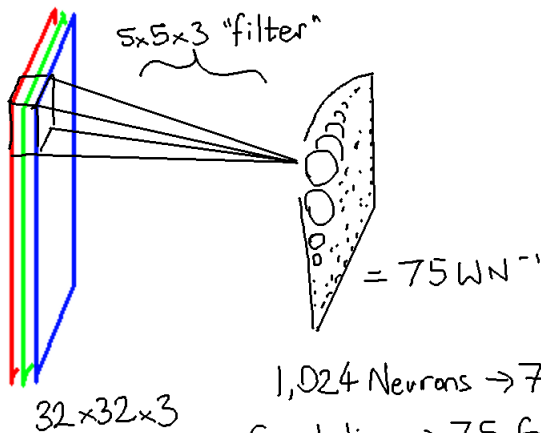
Gradient

Backprop

Activations

How to

Convolution



Convolution \rightarrow 75 for Layer

Local connectivity + convolution

Neural Networks

Declan Groves

What is

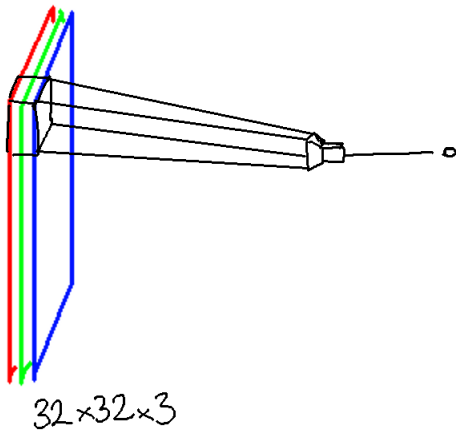
Gradient

Backprop

Activations

How to

Convolution



Local connectivity + convolution

Neural Networks

Declan Groves

What is

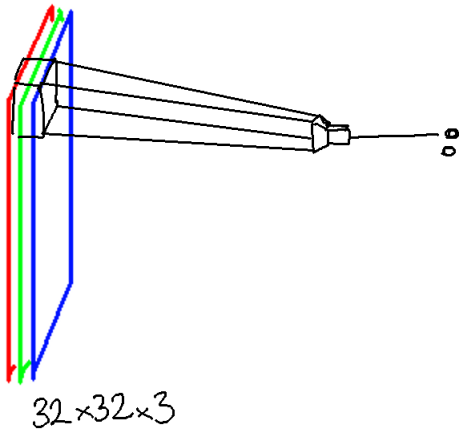
Gradient

Backprop

Activations

How to

Convolution



Local connectivity + convolution

Neural Networks

Declan Groves

What is

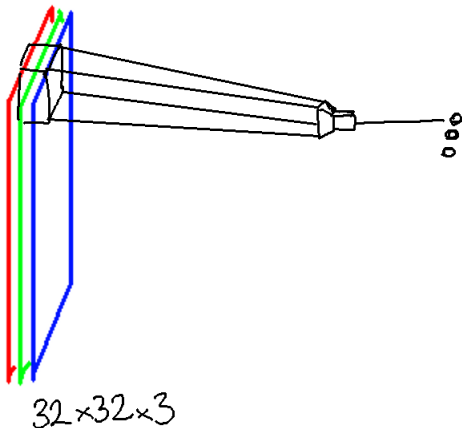
Gradient

Backprop

Activations

How to

Convolution



Local connectivity + convolution

Neural Networks

Declan Groves

What is

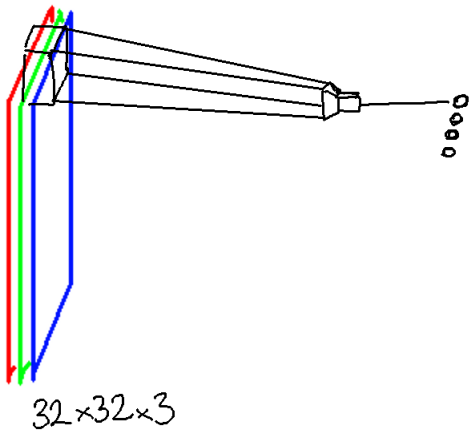
Gradient

Backprop

Activations

How to

Convolution



Local connectivity + convolution

Neural
Networks

Declan Groves

What is

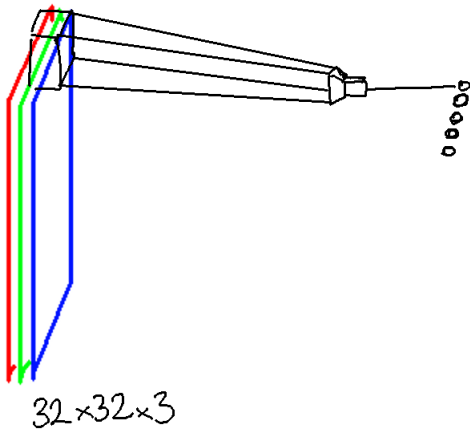
Gradient

Backprop

Activations

How to

Convolution



One filter forms a surface

Neural Networks

Declan Groves

What is

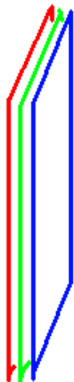
Gradient

Backprop

Activations

How to

Convolution



$32 \times 32 \times 3$



$32 \times 32 \times 1$

AlexNet's filters

Neural Networks

Declan Groves

What is

Gradient

Backprop

Activations

How to

Convolution

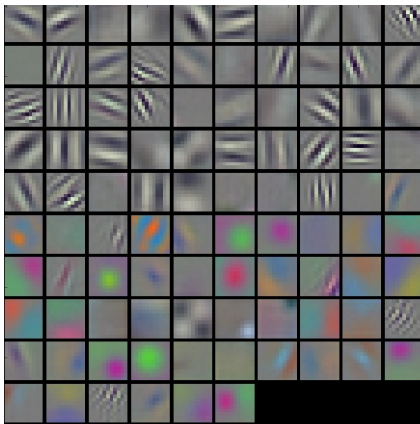


Image credit: Krizhevsky et al. 2012

Many filters form a volume

Neural
Networks

Declan Groves

What is

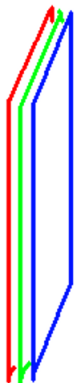
Gradient

Backprop

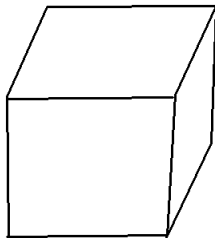
Activations

How to

Convolution



$32 \times 32 \times 3$



$32 \times 32 \times N_F$

Typical processing structure

Neural
Networks

Declan Groves

What is

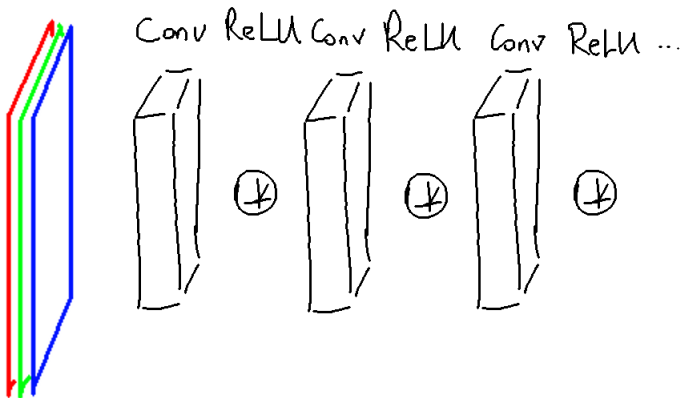
Gradient

Backprop

Activations

How to

Convolution



Pooling

Neural
Networks

Declan Groves

What is

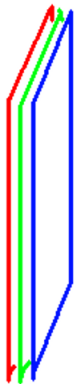
Gradient

Backprop

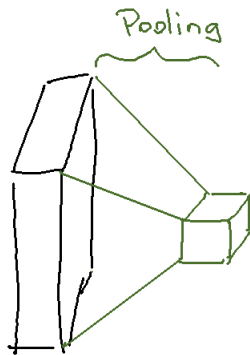
Activations

How to

Convolution



$\{\text{Conv}, \text{ReLU}\}_m$



$\{x, y, 3\} \rightarrow \{x, y, k\} \rightarrow \{\ll x, \ll y, k\}$

Pooling

Neural
Networks

Declan Groves

What is

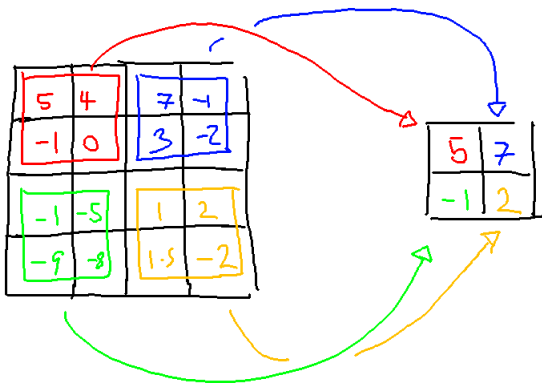
Gradient

Backprop

Activations

How to

Convolution



Typical structure of a CNN

Neural
Networks

Declan Groves

What is

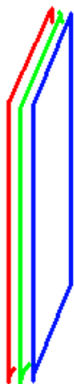
Gradient

Backprop

Activations

How to

Convolution



Standard Convnet

$$\{\{Conv, ReLU\}_m Pool\}_N \quad \{FC\}_K \rightarrow \hat{y}$$