

Neural Networks

Declan Groves

June 30 2016

Outline

Neural Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

- 1 What is a Neural Network?
- 2 Gradient Descent
- 3 Backprop
- 4 Activations
- 5 How to train your pet network
- 6 Convolutional Neural Networks

Overview

Neural Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

- Hypest ML
- Good at unstructured problems
- Suboptimal at structured problems

History

Neural Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

Backprop

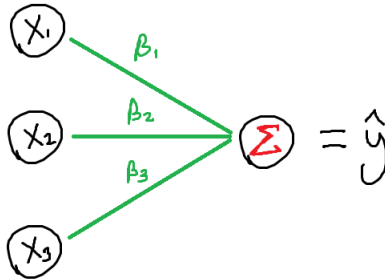
Activations

How to train
your pet
network

Convolutional
Neural
Networks

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

A graphical linear model



Neural
Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

A graphical linear model

Neural
Networks

Declan Groves

What is a
Neural
Network?

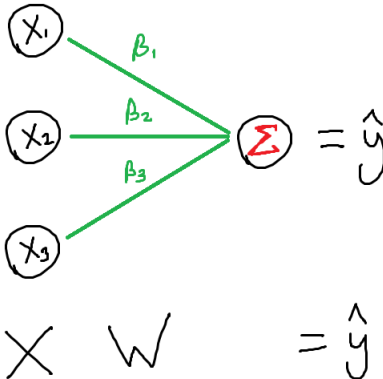
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



A graphical linear model

Neural Networks

Declan Groves

What is a Neural Network?

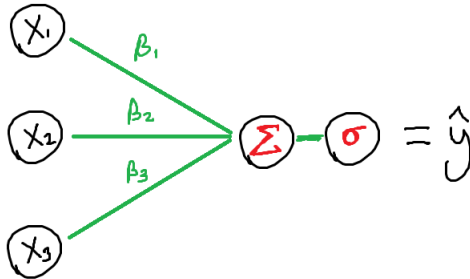
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

A graphical linear model

Neural
Networks

Declan Groves

What is a
Neural
Network?

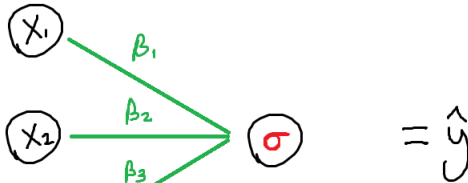
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



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

A simple neural network

Neural Networks

Declan Groves

What is a Neural Network?

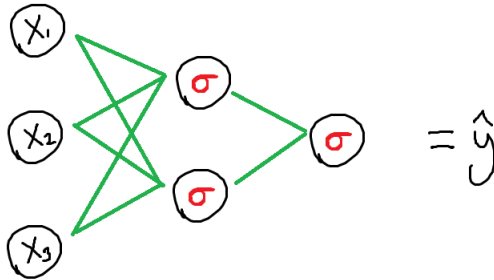
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



A simple neural network

Neural
Networks

Declan Groves

What is a
Neural
Network?

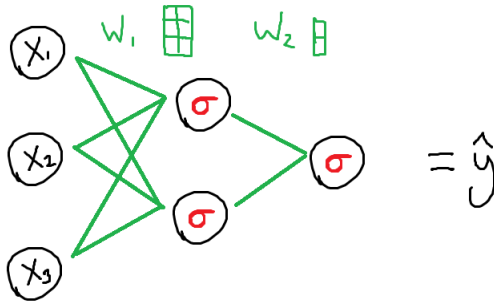
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



A simple neural network

Neural Networks

Declan Groves

What is a Neural Network?

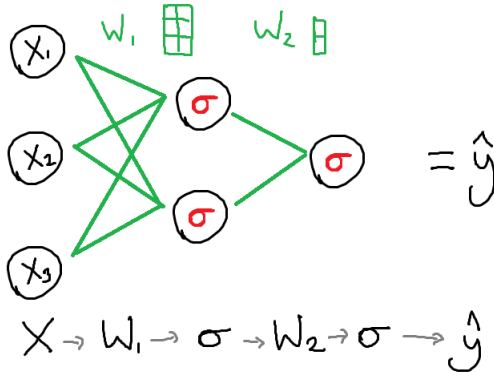
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



A simple neural network

Neural Networks

Declan Groves

What is a Neural Network?

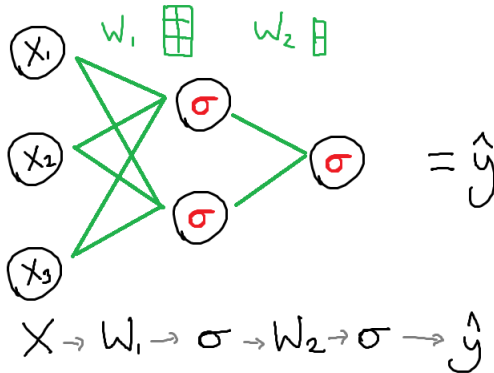
Gradient Descent

Backprop

Activations

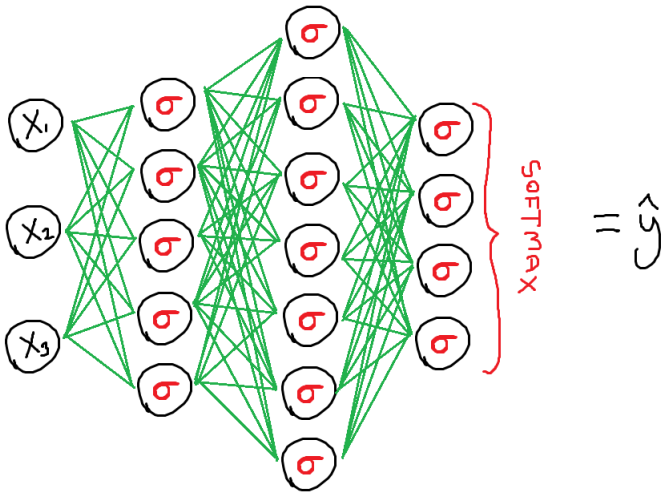
How to train your pet network

Convolutional Neural Networks



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

A multilayer multinomial classifier



Neural
Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

Gradient Descent

Neural Networks

Declan Groves

What is a Neural Network?

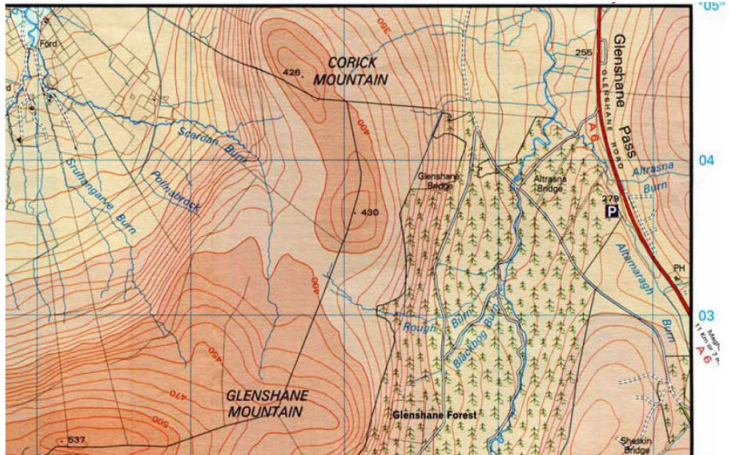
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Gradient Descent

Neural Networks

Declan Groves

What is a Neural Network?

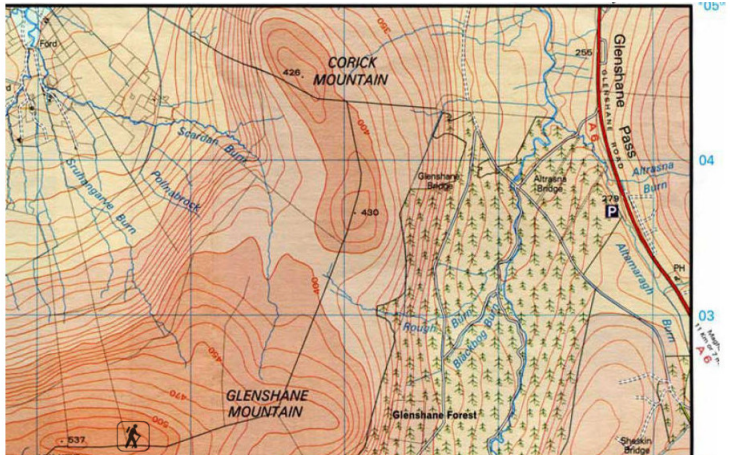
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Gradient Descent

Neural Networks

Declan Groves

What is a Neural Network?

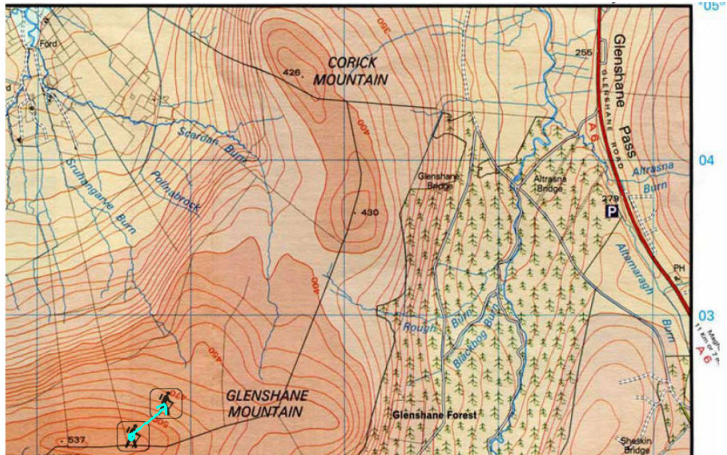
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

Gradient Descent

Neural Networks

Declan Groves

What is a Neural Network?

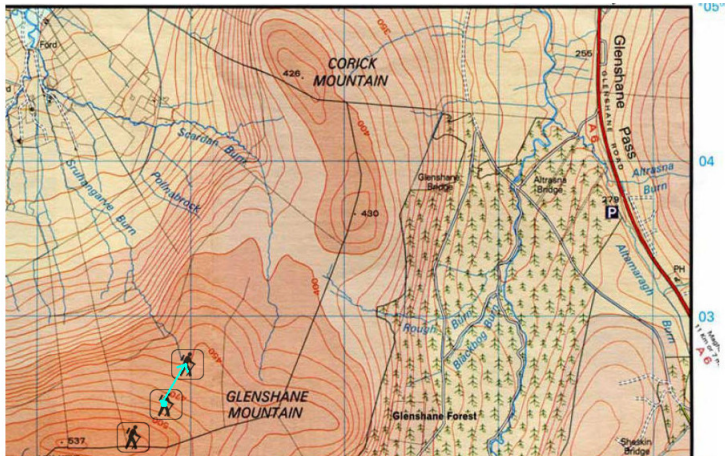
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

Gradient Descent

Neural
Networks

Declan Groves

What is a
Neural
Network?

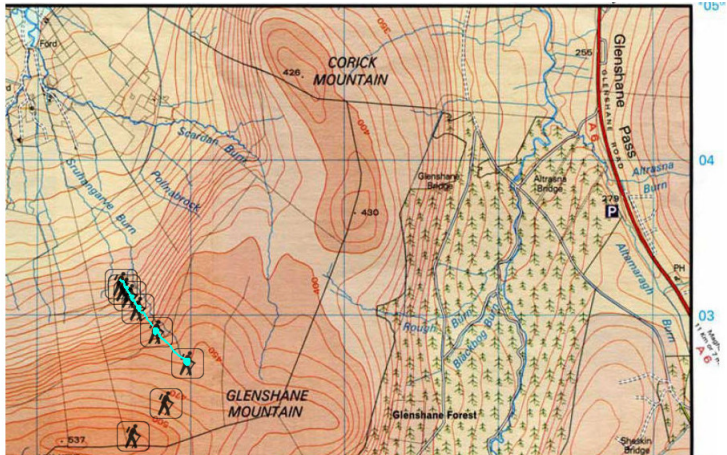
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



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

Gradient Descent and Momentum

Neural Networks

Declan Groves

What is a Neural Network?

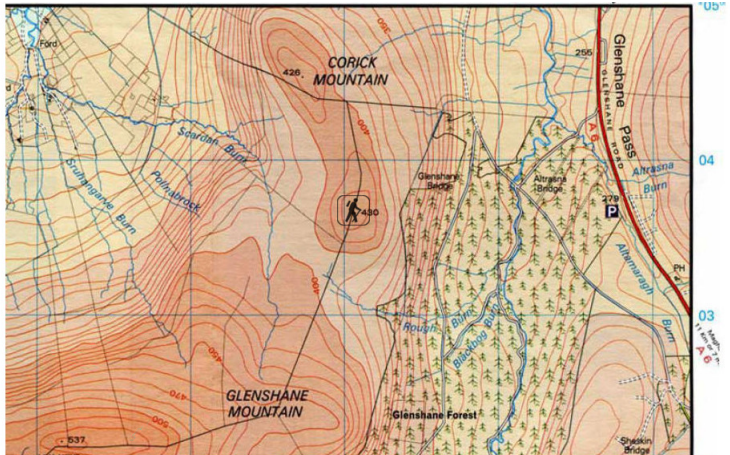
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Gradient Descent and Momentum

Neural Networks

Declan Groves

What is a Neural Network?

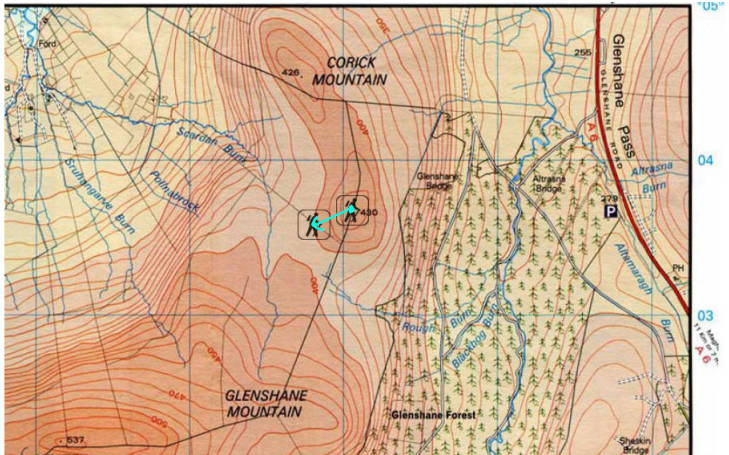
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Gradient Descent and Momentum

Neural
Networks

Declan Groves

What is a
Neural
Network?

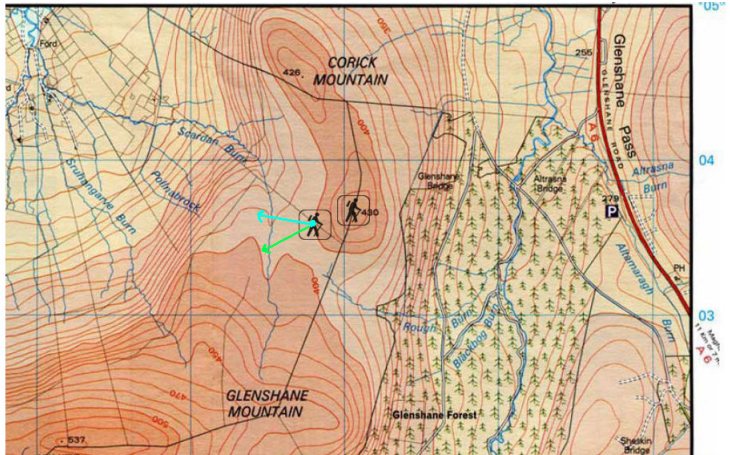
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



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

Gradient Descent and Momentum

Neural
Networks

Declan Groves

What is a
Neural
Network?

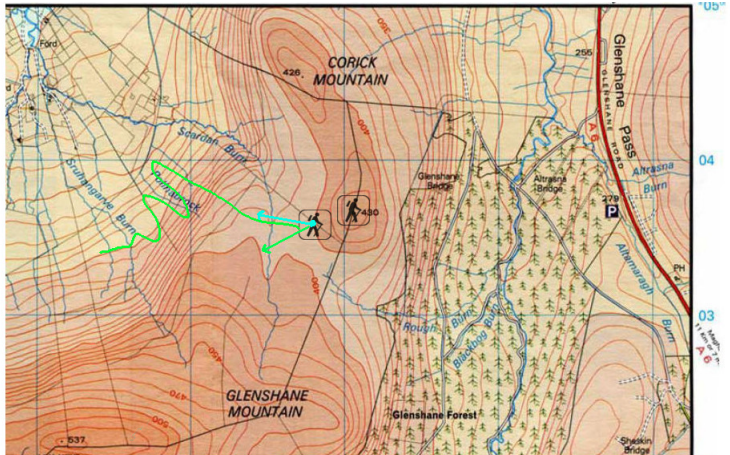
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



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

Ravines

Neural Networks

Declan Groves

What is a
Neural
Network?

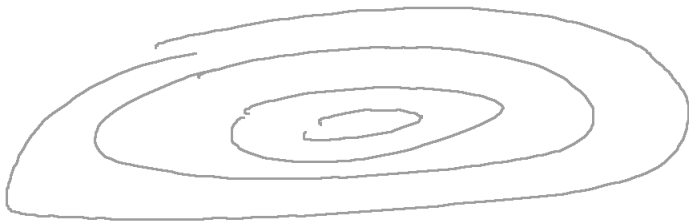
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Ravines - without momentum

Neural Networks

Declan Groves

What is a
Neural
Network?

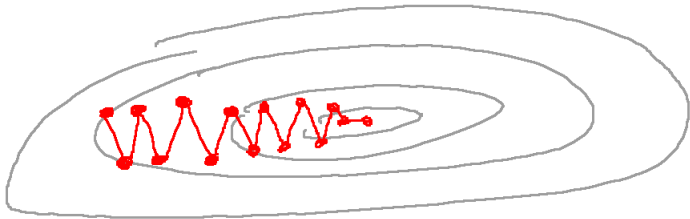
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Ravines - with momentum

Neural Networks

Declan Groves

What is a Neural Network?

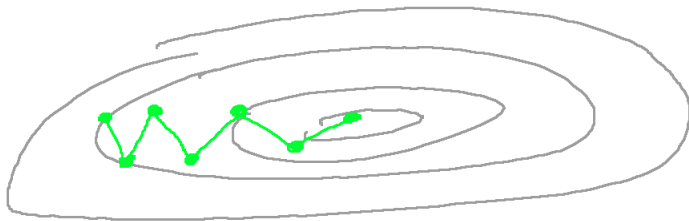
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Nesterov momentum

Neural
Networks

Declan Groves

What is a
Neural
Network?

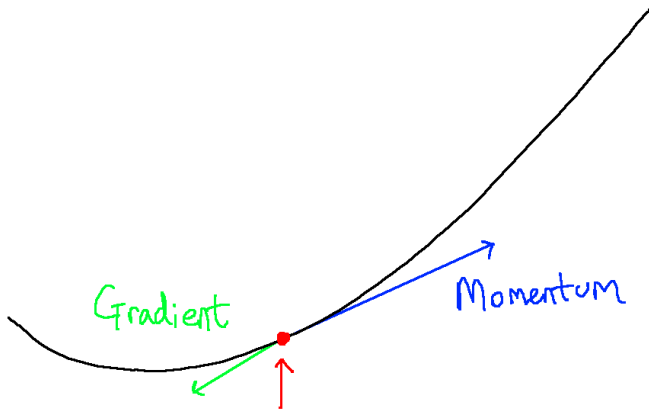
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Nesterov momentum

Neural
Networks

Declan Groves

What is a
Neural
Network?

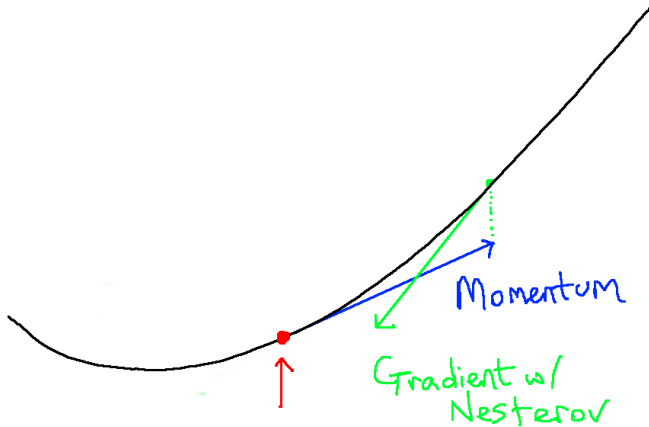
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Stochastic gradient descent

Neural
Networks

Declan Groves

What is a
Neural
Network?

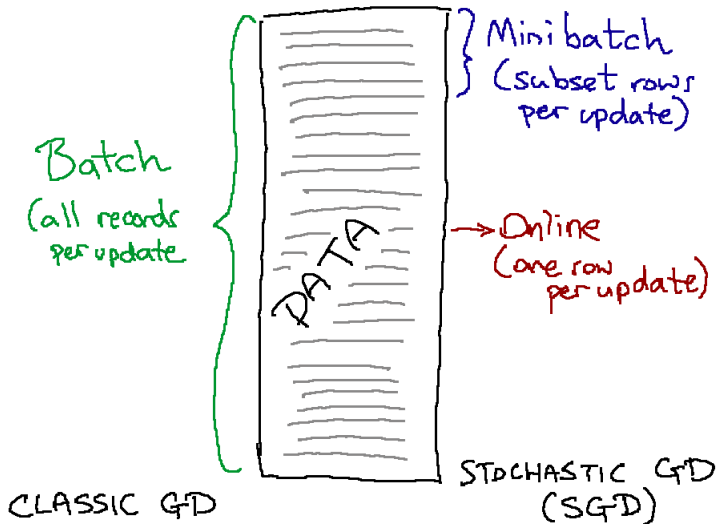
Gradient
Descent

Backprop

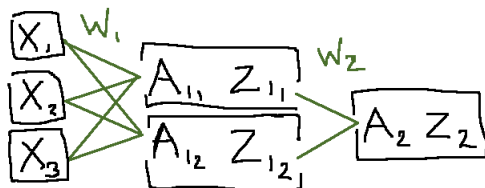
Activations

How to train
your pet
network

Convolutional
Neural
Networks



Forward pass



Neural
Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

Forward pass

Neural Networks

Declan Groves

What is a Neural Network?

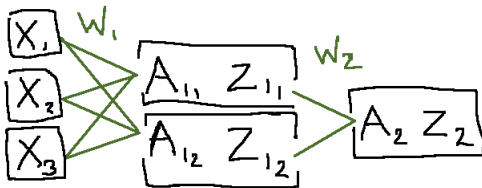
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

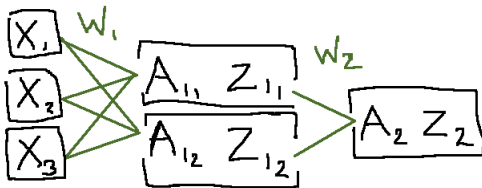
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

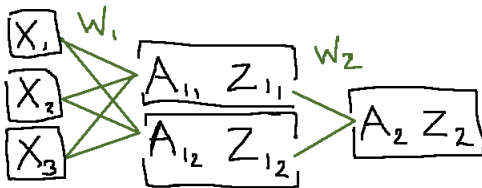
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

$$\partial J / \partial W_2 =$$

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

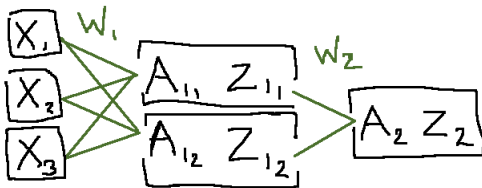
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

$$\frac{\partial J}{\partial w_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_2}$$

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

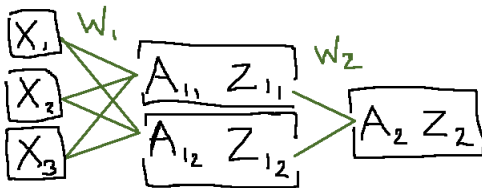
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

$$\frac{\partial J}{\partial w_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial A_2} \frac{\partial A_2}{\partial w_2}$$

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

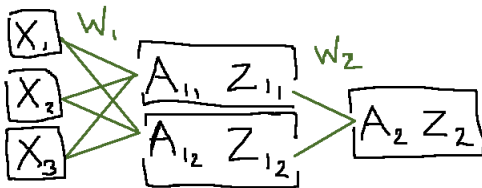
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

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

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

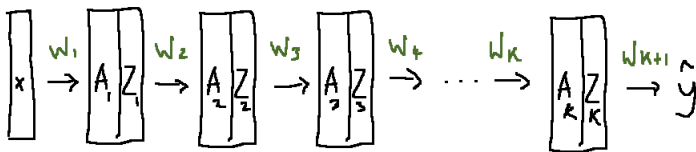
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



$$\partial J / \partial w_i =$$

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

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

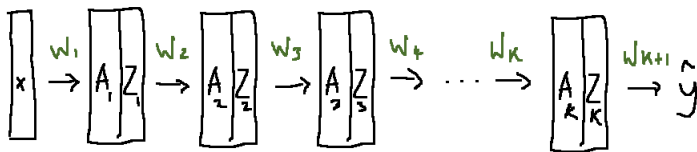
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



$$\partial J / \partial w_i =$$

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

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

!!

Backwards pass

Neural Networks

Declan Groves

What is a Neural Network?

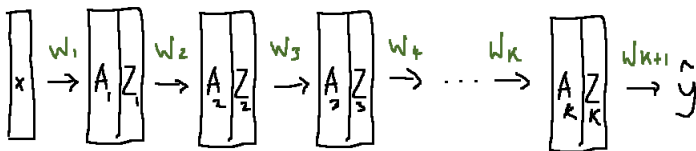
Gradient Descent

Backprop

Activations

How to train your pet network

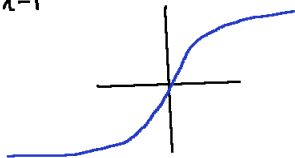
Convolutional Neural Networks



$$\frac{\partial J}{\partial w_i} =$$

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

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



Linear

Neural Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

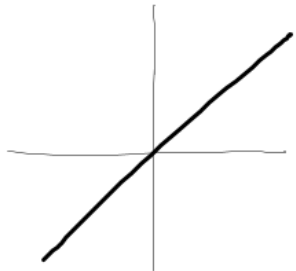
Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

■ Useless



Threshold

Neural Networks

Declan Groves

What is a Neural Network?

Gradient Descent

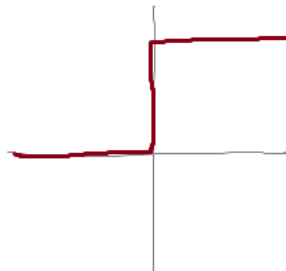
Backprop

Activations

How to train your pet network

Convolutional Neural Networks

- Similar to biological neuron
- No gradient



Sigmoid

Neural Networks

Declan Groves

What is a Neural Network?

Gradient Descent

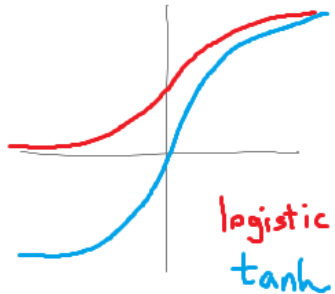
Backprop

Activations

How to train your pet network

Convolutional Neural Networks

- *tanh* preferred
- Gradients can vanish



ReLU

Neural Networks

Declan Groves

What is a Neural Network?

Gradient Descent

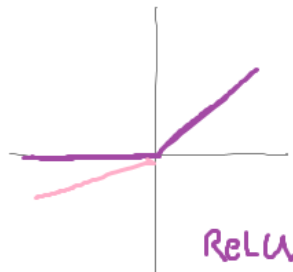
Backprop

Activations

How to train your pet network

Convolutional Neural Networks

- 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 a Neural Network?

Gradient Descent

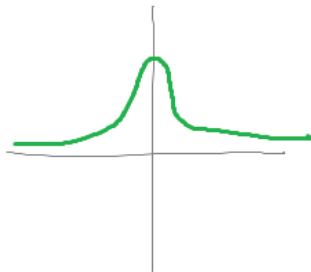
Backprop

Activations

How to train your pet network

Convolutional Neural Networks

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



How to train your pet network (generally)

Neural Networks

Declan Groves

What is a Neural Network?

Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks

- Use ReLU
- Use minibatch GD with Nesterov momentum
- ≤ 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 a
Neural
Network?

Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

- mxnet
- Keras

Convolutional Neural Networks

Neural
Networks

Declan Groves

What is a
Neural
Network?

Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks

- Image recognition killer

Conceptual structure

Neural Networks

Declan Groves

What is a Neural Network?

Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



MLP → too many weights!

Neural Networks

Declan Groves

What is a Neural Network?

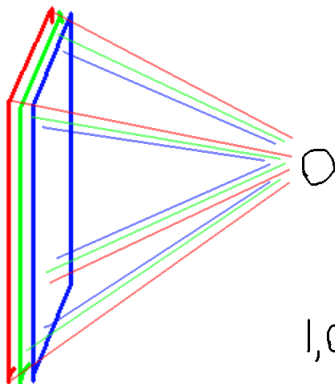
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



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

1,024 Neurons
→ $> 3M$ weights

Local connectivity

Neural Networks

Declan Groves

What is a
Neural
Network?

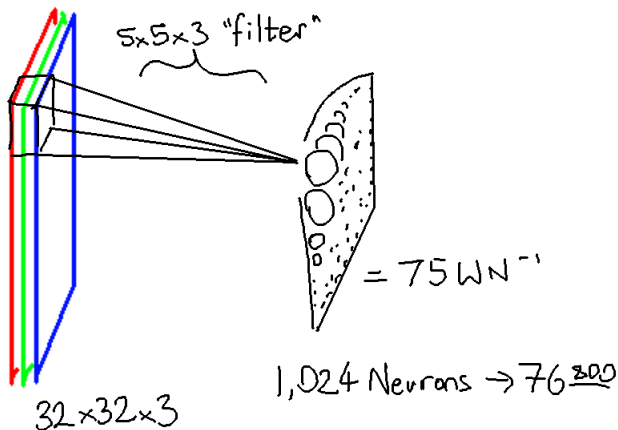
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Local connectivity + convolution

Neural Networks

Declan Groves

What is a Neural Network?

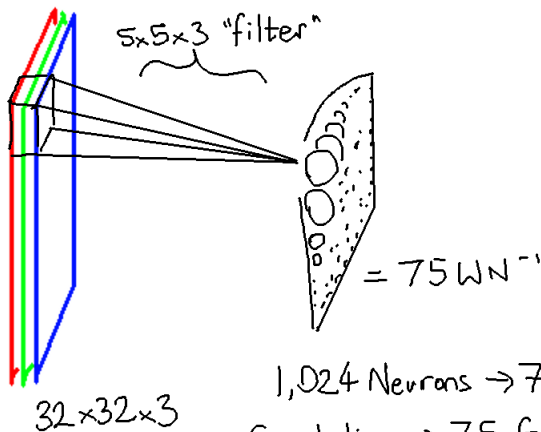
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Local connectivity + convolution

Neural Networks

Declan Groves

What is a Neural Network?

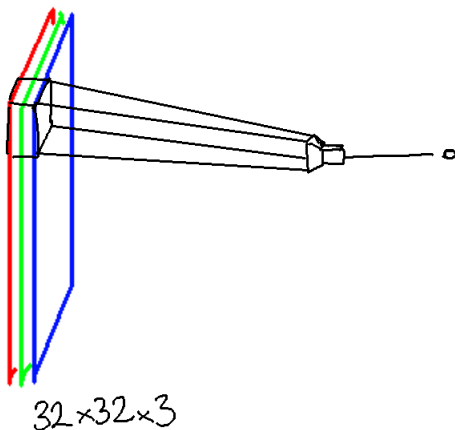
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Local connectivity + convolution

Neural Networks

Declan Groves

What is a
Neural
Network?

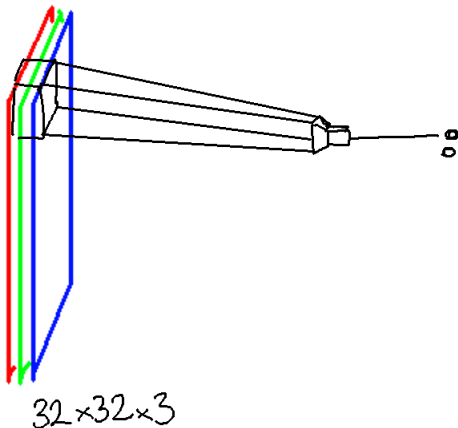
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Local connectivity + convolution

Neural Networks

Declan Groves

What is a
Neural
Network?

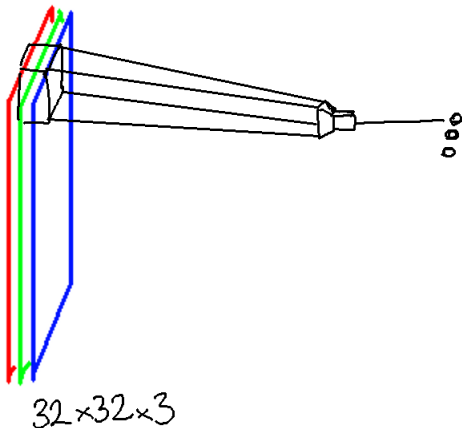
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Local connectivity + convolution

Neural Networks

Declan Groves

What is a
Neural
Network?

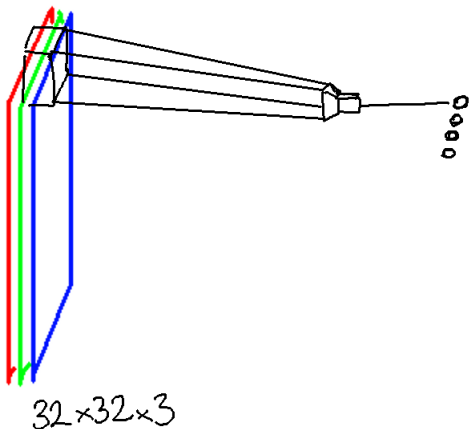
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Local connectivity + convolution

Neural Networks

Declan Groves

What is a
Neural
Network?

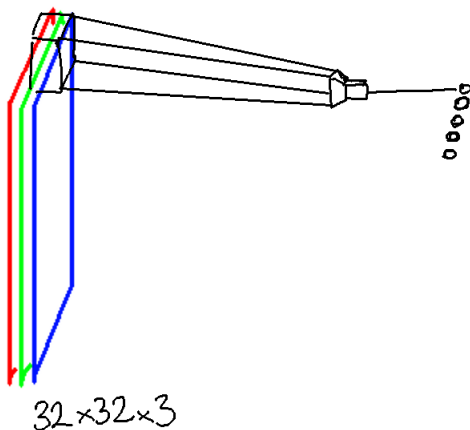
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



One filter forms a surface

Neural Networks

Declan Groves

What is a Neural Network?

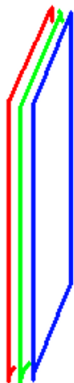
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



$32 \times 32 \times 3$



$32 \times 32 \times 1$

Many filters form a volume

Neural Networks

Declan Groves

What is a Neural Network?

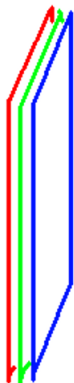
Gradient Descent

Backprop

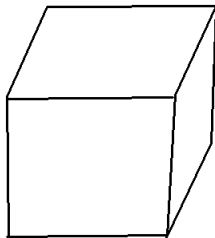
Activations

How to train your pet network

Convolutional Neural Networks



$$32 \times 32 \times 3$$



$$32 \times 32 \times N_F$$

Typical processing structure

Neural Networks

Declan Groves

What is a Neural Network?

Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



AlexNet's filters

Neural Networks

Declan Groves

What is a Neural Network?

Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks

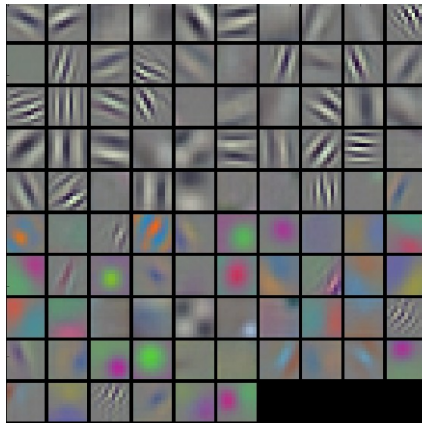


Image credit: Krizhevsky et al. 2012

Pooling

Neural Networks

Declan Groves

What is a Neural Network?

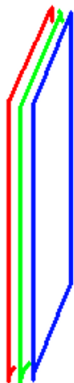
Gradient Descent

Backprop

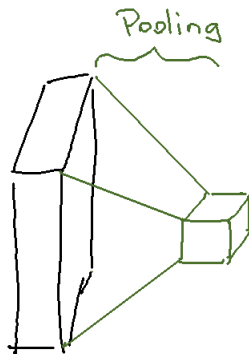
Activations

How to train your pet network

Convolutional Neural Networks



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



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

Pooling

Neural Networks

Declan Groves

What is a Neural Network?

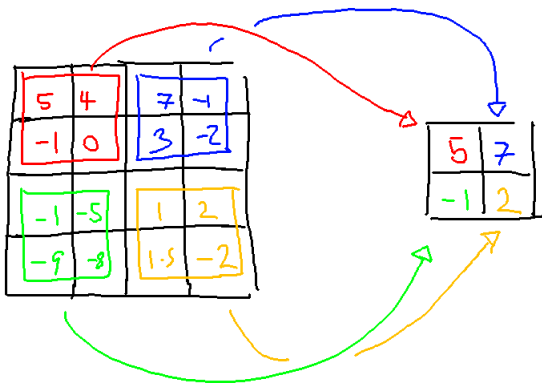
Gradient Descent

Backprop

Activations

How to train your pet network

Convolutional Neural Networks



Typical structure of a CNN

Neural
Networks

Declan Groves

What is a
Neural
Network?

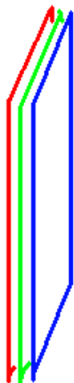
Gradient
Descent

Backprop

Activations

How to train
your pet
network

Convolutional
Neural
Networks



Standard Convnet

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