Neural Networks

Declan Grove

What i

----

Activation

Convolution

### **Neural Networks**

Declan Groves

June 30 2016

### Outline

#### Neural Networks

Declan Groves

What i

Gradien

Backproi

How to

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

#### Overview

#### Neural Networks

Declan Groves

What

Gradieni

Баскріо

Activati

How to

- 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

Gradien

Backpro

Activation

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

Neural Networks

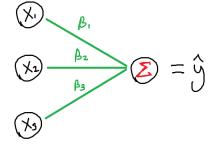
Declan Groves

What is

Gradien

Баскріо

Activation



Neural Networks

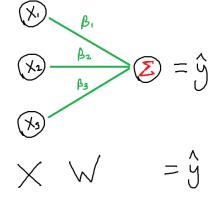
Declan Groves

What is

Gradien

Backpro

Activation



#### Neural Networks

Declan Groves

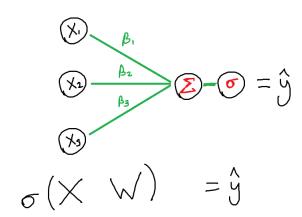
What is

Gradien

Backpro

Activation

How to



Neural Networks

Declan Groves

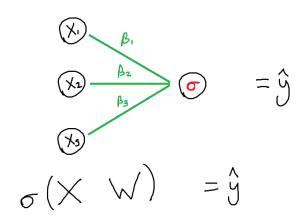
What is

C-----

Packers

A -+:..-+:--

How to



#### Neural Networks

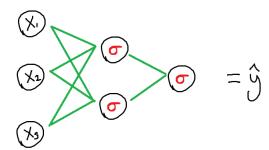
Declan Groves

#### What is

Gradien

Backpro

Activation



Neural Networks

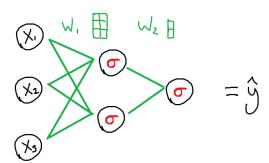
Declan Groves

What is

Gradier

васкрго

Activation



Neural Networks

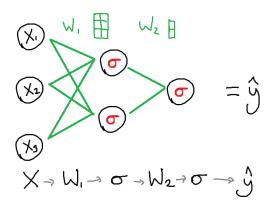
Declan Groves

What is

Gradien

Dackpro

Activation



Neural Networks

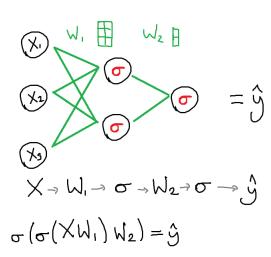
Declan Groves

What is

Cuadian

Backpro

Activation



# A multilayer multinomial classifier

Neural Networks

Declan Groves

What is

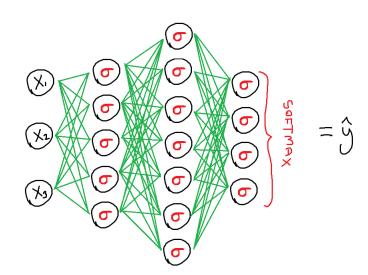
\_ ...

ь.

A ... ..

ACTIVATIO

How to



Neural Networks

Declan Groves

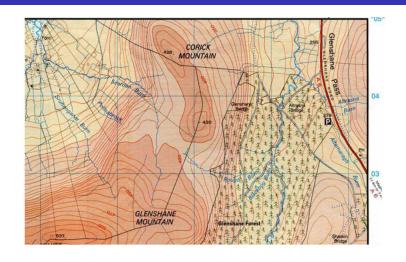
vviiat i

Gradient

Backpro

Activation

6 10



Neural Networks

Declan Groves

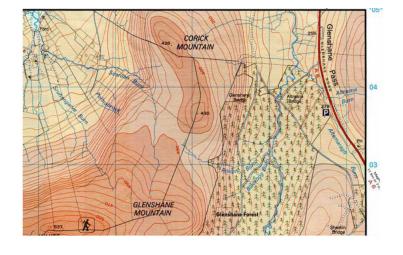
VVIIat I

Gradient

Backpro

Activation

Campalanta



Neural Networks

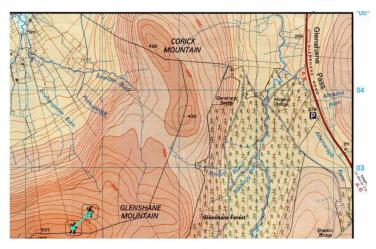
Declan Groves

vvnat i

Gradient

Backpro

Activation



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

Neural Networks

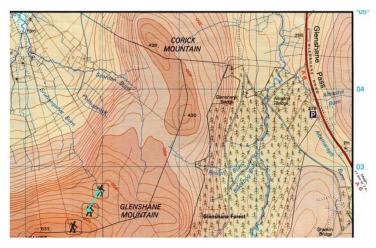
Declan Groves

What i

Gradient

Backpro

Activation



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

Neural Networks

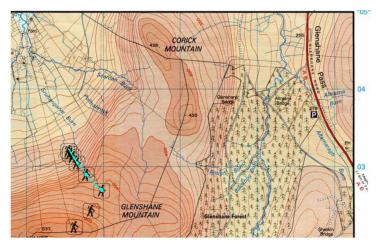
Declan Groves

vvnati

Gradient

Backpro

Activation



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

Neural Networks

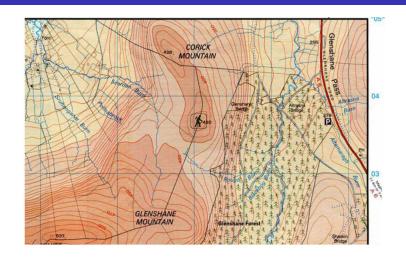
Declan Groves

What is

Gradient

Rackpro

A set set s



Neural Networks

Declan Groves

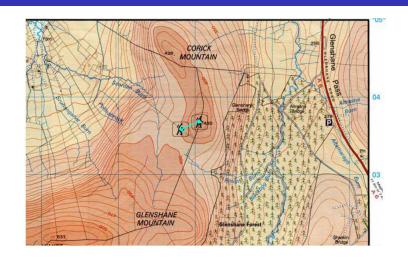
What is

Gradient

Rackpro

Activation

6 10



Neural Networks

Declan Groves

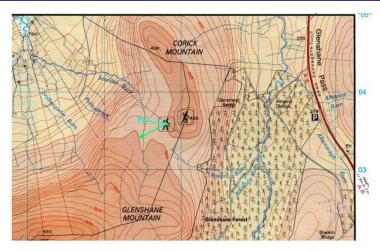
What i

Gradient

Backpro

Activation

, (01,741,0)



Neural Networks

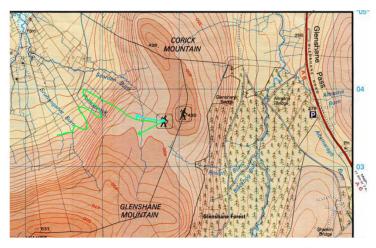
Declan Groves

What i

Gradient

Backpro

Activation



### Ravines

Neural Networks

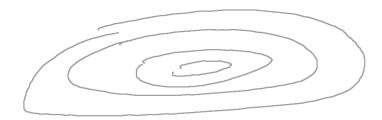
Declan Groves

What i

Gradient

Backprop

Activation



### Ravines - without momentum

Neural Networks

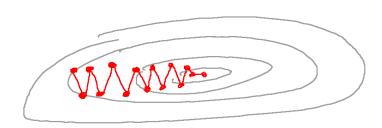
Declan Groves

What is

Gradient

Daaliana

Activation



### Ravines - with momentum

Neural Networks

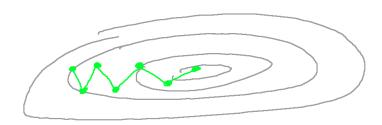
Declan Groves

What is

Gradient

Backprop

Activation



### Nesterov momentum

Neural Networks

Declan Groves

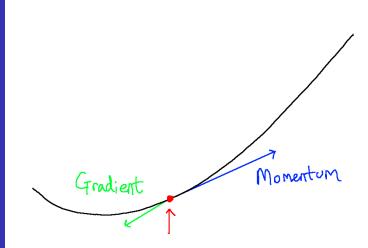
What i

Gradient

\_ .

Activation

How to



### Nesterov momentum

Neural Networks

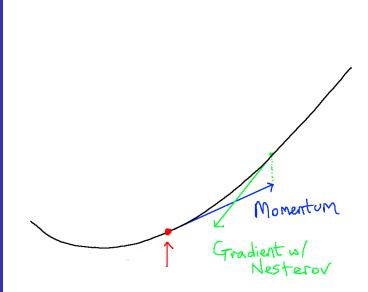
Declan Groves

What i

Gradient

\_ .

Activation



# Stochastic gradient descent

Neural Networks

Declan Groves

What is

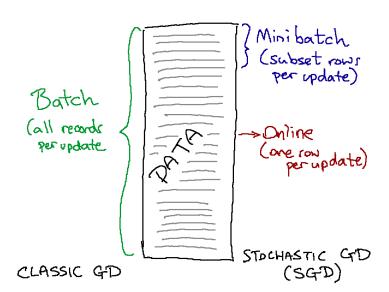
Gradient

Backproi

A - 11 - - 11 - - -

Activation

How to



# Forward pass

Neural Networks

Declan Groves

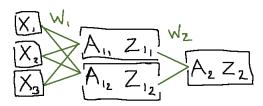
What i

Gradient

Backprop

A ...

How to



# Forward pass

Neural Networks

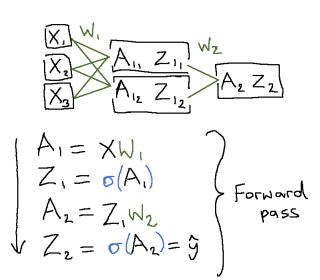
Declan Groves

What i

Gradien

Backprop

Activation



Neural Networks

Declan Groves

What i

Gradier

Backprop

Activation

$$X_1$$
 $X_2$ 
 $X_3$ 
 $A_1$ 
 $A_1$ 
 $A_2$ 
 $A_2$ 
 $A_2$ 
 $A_2$ 
 $A_3$ 
 $A_4$ 
 $A_5$ 
 $A_5$ 

Neural Networks

Declan Groves

What i

Gradien

Backprop

Activation

$$\begin{array}{c|c} X_1 & W_1 \\ \hline X_2 & A_{11} & Z_{11} \\ \hline X_3 & A_{12} & Z_{12} \\ \hline \end{array}$$

$$\begin{array}{c|c} A_1 & Z_{12} \\ \hline A_2 & Z_2 \\ \hline \end{array}$$

$$\begin{array}{c|c} A_2 & Z_2 \\ \hline \end{array}$$

$$\begin{array}{c|c} J = \text{err}(y, \hat{y})_1 & J \leftarrow \frac{1}{2}(y - \hat{y})^2 \\ \hline \partial J/\partial U_2 & = \end{array}$$

Neural Networks

Backprop

$$\begin{array}{c|c} X_1 & W_1 \\ \hline X_2 & A_{12} & Z_{12} \\ \hline A_{12} & Z_{12} \\ \hline A_{2} & Z_{2} \\ \hline D J/\partial U_2 = \partial J/\partial \hat{y} & \partial \hat{y}/\partial W_2 \\ \end{array}$$

Neural Networks

Declan Groves

What i

Gradien

Backprop

Activation

$$\begin{array}{c|c} X_1 & W_1 \\ \hline X_2 & A_{11} & Z_{12} \\ \hline X_3 & A_{12} & Z_{12} \\ \hline \end{array}$$

$$\begin{array}{c|c} A_1 & Z_{12} & A_2 & Z_2 \\ \hline \end{array}$$

$$\begin{array}{c|c} J = err(y, \hat{y}), J \leftarrow \frac{1}{2}(y - \hat{y})^2 \\ \partial J/\partial W_2 &= \partial J/\partial \hat{y} & \partial \hat{y}/\partial A_2 & \partial A_2/\partial W_2 \end{array}$$

Neural Networks

Declan Groves

What i

Gradien

Backprop

Activation

$$\begin{array}{c|c}
X_1 & W_1 \\
\hline
X_2 & A_1, Z_{1_1} & W_2 \\
\hline
A_1, Z_{1_2} & A_2 Z_2
\end{array}$$

$$\begin{array}{c|c}
T = err(y, \hat{y}), \quad J \leftarrow \frac{1}{2}(y - \hat{y})^2 \\
\partial J/\partial W_2 &= \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial A_2} \frac{\partial A_2}{\partial W_2} \\
&= (y - \hat{y}) \quad \sigma'(A_2) \quad Z_1
\end{array}$$

#### Neural Networks

Declan Groves

What

Gradien

Backprop

Activation

$$\frac{\partial J}{\partial \hat{y}} = \frac{\partial J}{\partial \hat{y}} \sigma'(A_{k}) W_{k} \sigma'(A_{k-1}) W_{k-1} \cdots W_{i+1} \sigma'(A_{k}) Z_{i-1}$$

### Backwards pass

#### Neural Networks

Declan Groves

What

Gradien

Backprop

Activation

$$\partial J/\partial W_{i} =$$

$$(\partial J/\partial \hat{y}) \sigma'(A_{k}) W_{k} \sigma'(A_{k-1}) W_{k-1} ...$$

$$... W_{i+1} \sigma'(A_{k}) Z_{k-1}$$

### Backwards pass

#### Neural Networks

Declan Groves

What

Gradien

Backprop

Activation

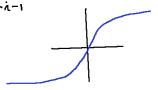
, (01,101,0)

Campalitation

$$\partial J/\partial W_{i} =$$

$$(\partial J/\partial \dot{y}) \underline{\sigma}'(A_{k}) W_{k} \underline{\sigma}'(A_{k-1}) W_{k-1} ...$$

$$... W_{i+1} \underline{\sigma}'(A_{k}) Z_{k-1}$$



#### Linear

Neural Networks

Declan Groves

What i

C == d:==

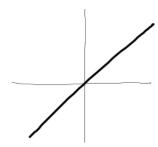
Backpro

Activations

. .

Convolution

Useless



#### Threshold

#### Neural Networks

Declan Groves

What i

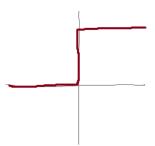
C-- J:-.

Backpro

Activations

How to

- Similar to biological neuron
- No gradient



## Sigmoid

Neural Networks

Declan Groves

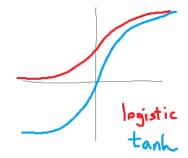
What i

Gradient

васкрго

Activations

- tanh preferred
- Gradients can vanish



#### ReLU

#### Neural Networks

Declan Groves

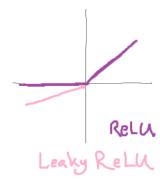
What i

Gradient

Backpro

Activations

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



#### Radial basis functions

#### Neural Networks

Declan Groves

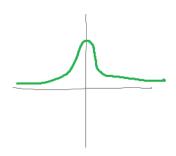
What

Gradient

Backpro

Activations

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



### Six simple rules for training your neural network

#### Neural Networks

How to

- Use Rel U
- Use minibatch GD with Nesterov momentum ( $\approx 0.9$ )
- < 3 layers (unless convolutional)</p>
- Use dropout ( $\approx 0.2$  input layer,  $\approx 0.5$  else)
- Prefer wider with L2 over smaller
- Be careful with weight initialization!

#### Libraries for newbs

Neural Networks

Declan Groves

What i

Gradien

Backpror

Activatio

How to

- mxnet
- Keras

#### Convolutional Neural Networks

Neural Networks

Declan Groves

What

Gradien

Backprop

/ tetivat

How to

Convolution

■ Image recognition killer

### Conceptual structure

Neural Networks

Declan Groves

What is

O. da.c..

Backpro

Activation

, (01,101,0)

## $\mathsf{MLP} \to \mathsf{too}$ many weights!

Neural Networks

Declan Groves

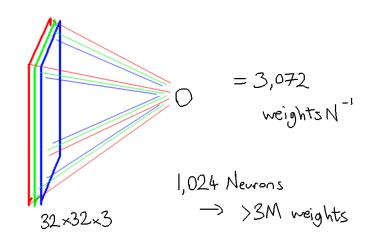
What is

C-----

Packero

Activatio

How to



## Local connectivity

Neural Networks

Declan Grove

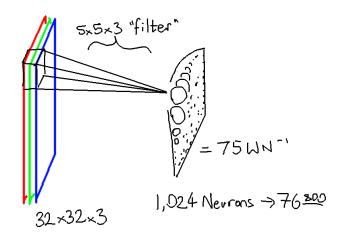
What is

Gradient

Backpro

Activatio

110W LO



Neural Networks

Declan Grove

What is

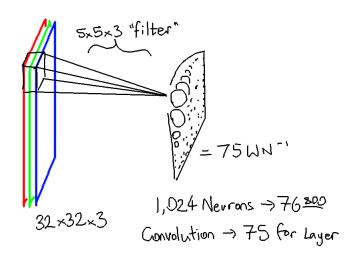
Gradien

Backpro

A -4:...4:-

/ (Ctivatio

How to



Neural Networks

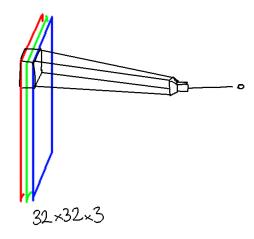
Declan Groves

What i

Gradien

Backpro

Activation



Neural Networks

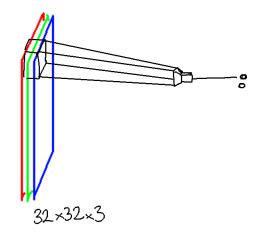
Declan Groves

What i

Gradien

Backpro

Activatio



Neural Networks

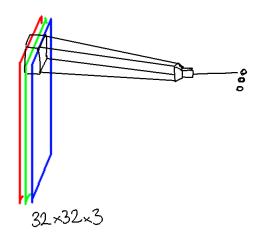
Declan Groves

What is

Gradien

Backpro

Activatio



## Local connectivity + convolution

Neural Networks

Declan Groves

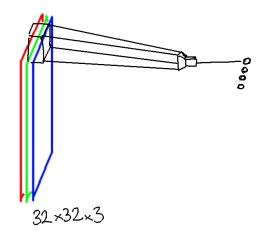
What is

\_\_\_\_\_

васкрго

Activatio

...



Neural Networks

Declan Groves

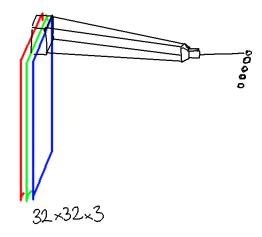
What is

Gradieni

Баскргс

Activati

How to



### One filter forms a surface

Neural Networks

Declan Groves

What is

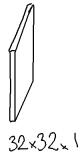
Gradien

Backpro

Activati

How to





#### AlexNet's filters

Neural Networks

Declan Groves

What is

C----

Rackpro

Activatio

How to

Convolution

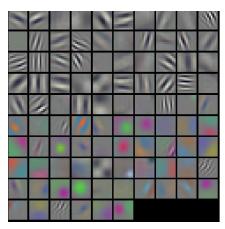


Image credit: Krizhevsky et al. 2012

# Many filters form a volume

Neural Networks

Declan Groves

What is

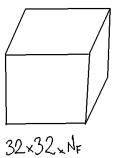
Gradier

Backpro

Activation

How to





## Typical processing structure

Neural Networks

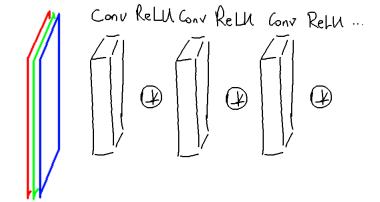
Declan Groves

What is

Gradieni

Backpro

Activation



## **Pooling**

Neural Networks

Declan Groves

What is

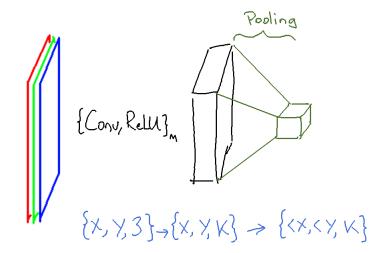
Gradier

Backpro

Activati

. .

11000 10



# **Pooling**

Neural Networks

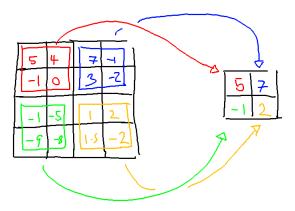
Declan Grove

What is

Gradieni

Backpro

Activatio



### Typical structure of a CNN

Neural Networks

Declan Groves

What is

Gradien

Backpro

A -4:..-4:--

Activation

