NEURAL NETWORKS APPLIED

David Kauchak
CS159 – Fall 2024

1

Admin

Assignment 5a due tonight

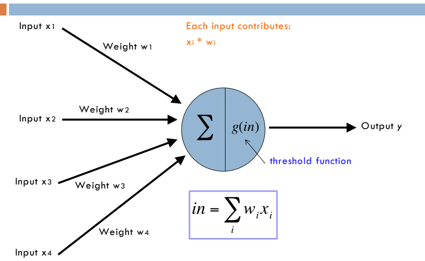
Assignment 5b due next Wednesday

Grading update

Schedule for the rest of the semester mostly up to date

2

A Single Neuron/Perceptron

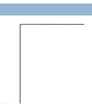


3

Activation functions

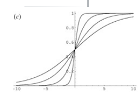
hard threshold:

$$g(in) = \begin{cases} 1 & if\ in \geq T \\ 0 & otherwise \end{cases}$$

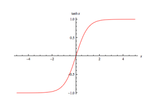


sigmoid

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



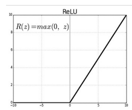
tanh x



4

Many other activation functions

Rectified Linear Unit

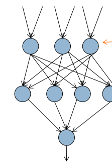


Softmax (for probabilities)

5

Neural network

inputs



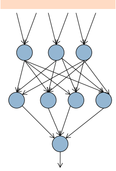
Individual perceptrons/neurons

6

Neural network

inputs

some inputs are provided/entered

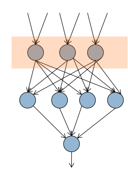


7

Neural network

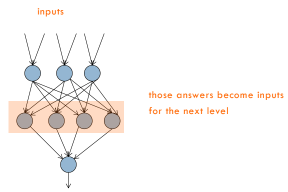
inputs

each perceptron computes and calculates an answer



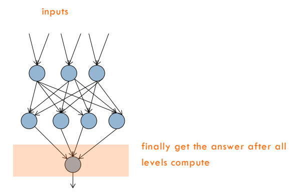
8

Neural network



9

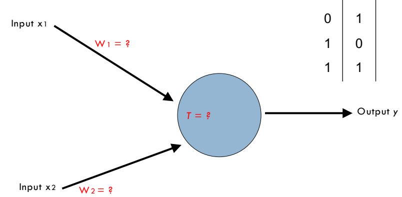
Neural network



10

XOR

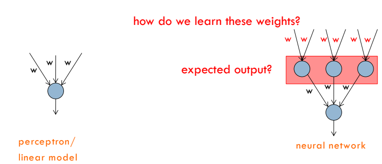
x_1	x_2	x_1 **xor** x_2
0	0	0
0	1	1
1	0	1
1	1	0



11

Learning in multilayer networks

Challenge: for multilayer networks, we don't know what the expected output/error is for the internal nodes!



12

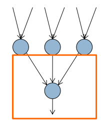
Backpropagation: intuition

Gradient descent method for learning weights by optimizing a loss function

1. calculate output of all nodes
2. calculate the weights for the output layer based on the error
3. “backpropagate” errors through hidden layers

13

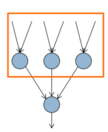
Backpropagation: intuition



We can calculate the actual error here

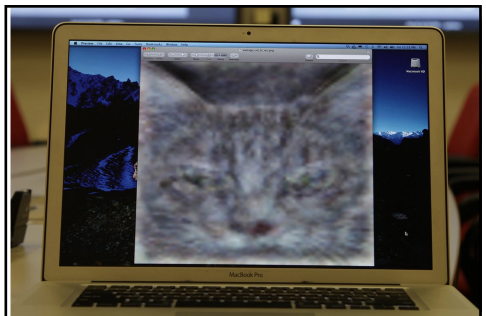
14

Backpropagation: intuition



Key idea: propagate the error back to this layer

15



Trained a NN with 1B connections on 10M snapshots from youtube on 16,000 processors

16

http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html

17

Deep learning

WIKIPEDIA

Deep learning is a branch of machine learning based on a set of algorithms that attempt to model high level abstractions in data by using a deep graph with multiple processing layers, composed of multiple linear and non-linear transformations.

Deep learning is part of a broader family of machine learning methods based on learning representations of data.

18

Deep learning

Key: learning better features that abstract from the "raw" data

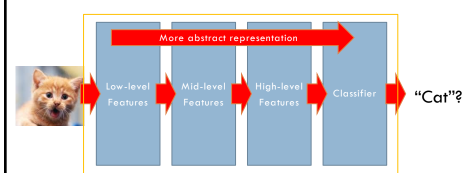
Using learned feature representations based on large amounts of data, generally unsupervised

Using classifiers with multiple layers of learning

19

Deep learning

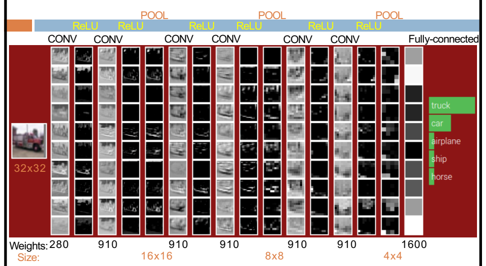
- Train *multiple layers* of features/abstractions from data.
- Try to discover *representation* that makes decisions easy.



Deep Learning: train layers of features so that classifier works well.

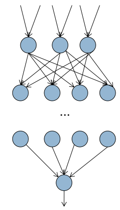
Slide adapted from: Adam Coates

20



21

Deep learning for neural networks



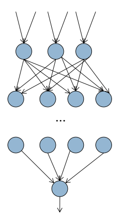
Traditional NN models: 1-2 hidden layers

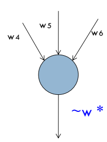
Deep learning NN models: 3+ hidden layers

22

Challenges

What makes "deep learning" hard for NNs?





Modified errors tend to get diluted as they get combined with many layers of weight corrections

23

Deep learning

Growing field

Driven by:

- Increase in data availability
- Increase in computational power
- Parallelizability of many of the algorithms

Involves more than just neural networks (though, they're a very popular model)

24

word2vec

How many people have heard of it?

What is it?

25

Word representations generalized

Project words into a multi-dimensional “meaning” space

word ➡ $[x_1, x_2, \ldots, x_d]$

What is our projection for assignment 5?

26

Word representations generalized

Project words into a multi-dimensional “meaning” space

word ➡ $[w_1, w_2, \ldots, w_d]$

Each dimension is the co-occurrence of *word* with w_i

27

Word representations

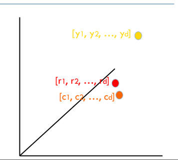
Project words into a multi-dimensional “meaning” space

word ➡ $[x_1, x_2, \ldots, x_d]$

red ➡ $[r_1, r_2, \ldots, r_d]$

crimson ➡ $[c_1, c_2, \ldots, c_d]$

yellow ➡ $[y_1, y_2, \ldots, y_d]$



28

Word representations

Project words into a multi-dimensional "meaning" space

word ➡ $[x_1, x_2, \ldots, x_d]$

The idea of word representations is not new:
- Co-occurrence matrices
- Latent Semantic Analysis (LSA)

New idea: learn word representation using a task-driven approach

29

A prediction problem

I like to eat bananas with cream cheese

Given a context of words

Predict what words are likely to occur in that context

30

A prediction problem

Given text, can generate lots of examples:

I like to eat bananas with cream cheese

input	prediction
___ like to eat	I
I ___ to eat bananas	like
I like ___ eat bananas with	to
I like to ___ bananas with cream	eat
…	…

31

A prediction problem

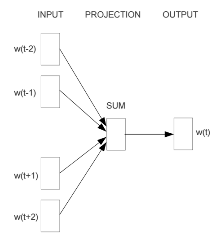
Use data like this to learn a distribution:

$$p(word \mid context)$$

$$p(w_i \mid \underbrace{w_{i-2} w_{i-1}}_{\text{words before}} \underbrace{w_{i+1} w_{i+2}}_{\text{words after}})$$

32

Train a neural network on this problem

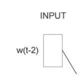


https://arxiv.org/pdf/1301.3781v3.pdf

33

Encoding words

How can we input a "word" into a network?

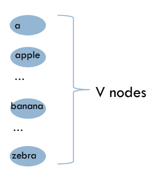


34

"One-hot" encoding

For a vocabulary of V words, have V input nodes

All inputs are 0 except for the one corresponding to the word

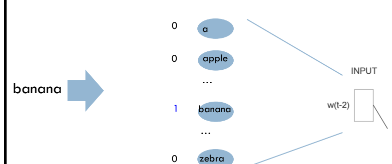


35

"One-hot" encoding

For a vocabulary of V words, have V input nodes

All inputs are 0 except the for the one corresponding to the word



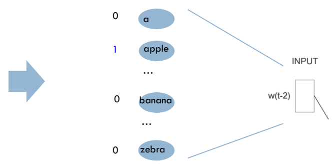
36

"One-hot" encoding

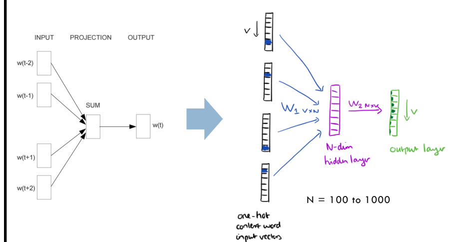
For a vocabulary of V words, have V input nodes

All inputs are 0 except the for the one corresponding to the word

apple



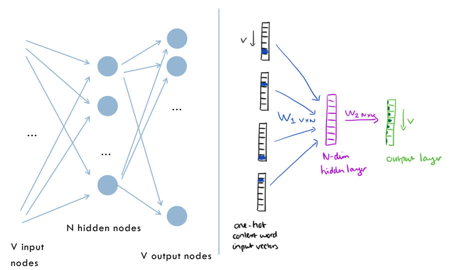
37



https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/

38

Another view



39

Training: backpropagation

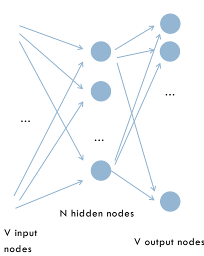
___ like to eat

I ___ to eat bananas

I like ___ eat bananas with

I like to ___ bananas with cream

...



I

like

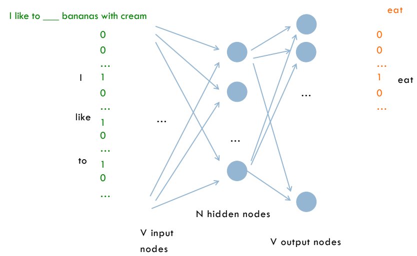
to

eat

...

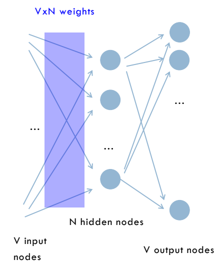
40

Training: backpropagation



41

Word representation



The weights for each word provide an N dimensional mapping of the word

Words that predict similarly should have similar weights

Why does this work?

42

Results

vector(word1) – vector(word2) = vector(word3) - X

word1 is to word2 as word3 is to X

Type of relationship	Word Pair 1		Word Pair 2	
Common capital city	Athens	Greece	Oslo	Norway
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe
Currency	Angola	kwanza	Iran	rial
City-in-state	Chicago	Illinois	Stockton	California
Man-Woman	brother	sister	grandson	granddaughter

43

Results

vector(word1) – vector(word2) = vector(word3) - X

word1 is to word2 as word3 is to X

Type of relationship	Word Pair 1		Word Pair 2	
Adjective to adverb	apparent	apparently	rapid	rapidly
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Superlative	easy	easiest	lucky	luckiest
Present Participle	think	thinking	read	reading
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars
Plural verbs	work	works	speak	speaks

44

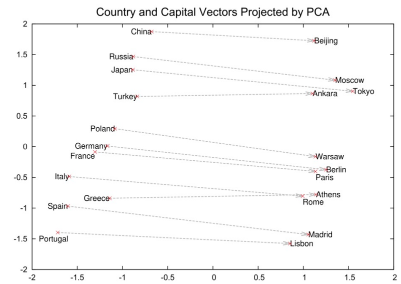
Results

vector(word1) – vector(word2) = vector(word3) - X

word1 is to word2 as word3 is to X

Newspapers			
New York San Jose	New York Times San Jose Mercury News	Baltimore Cincinnati	Baltimore Sun Cincinnati Enquirer
NHL Teams			
Boston Phoenix	Boston Bruins Phoenix Coyotes	Montreal Nashville	Montreal Canadiens Nashville Predators
NBA Teams			
Detroit Oakland	Detroit Pistons Golden State Warriors	Toronto Memphis	Toronto Raptors Memphis Grizzlies
Airlines			
Austria Belgium	Austrian Airlines Brussels Airlines	Spain Greece	Spainair Aegean Airlines
Company executives			
Steve Ballmer Samuel J. Palmisano	Microsoft IBM	Larry Page Werner Vogels	Google Amazon

45



2-Dimensional projection of the N-dimensional space

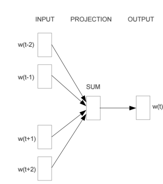
46

Visualized

https://projector.tensorflow.org/

47

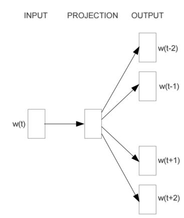
Continuous Bag Of Words



CBOW

48

Other models: skip-gram



49

word2vec

A model for learning word representations from large amounts of data

Has become a popular pre-processing step for learning a more robust feature representation

Models like word2vec have also been incorporated into other learning approaches (e.g. translation tasks)

50

word2vec resources

https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/

https://code.google.com/archive/p/word2vec/

https://deeplearning4j.org/word2vec

https://arxiv.org/pdf/1301.3781v3.pdf

51

Playing with word2vec

http://vectors.nlpl.eu/explore/embeddings/en/

52

10 minutes

https://www.youtube.com/watch?v=zI99IZvW7rE

53