Deep Learning for NLP

Younes Samih

Department of Computational Linguistics

November 8, 2017

Road Map

Prerequisites

- Python

Prerequisites

- Python
- Linear Algebra

Expectation from the Course

- you will be able to understand and

Expectation from the Course

- you will be able to understand and
- build neural network models using python

Expectation from the Course

- you will be able to understand and
- build neural network models using python
- you will be able to solve some NLP Problems

Software

- Python (`https://www.python.org/`)

Software

- Python (`https://www.python.org/`)
- Numpy (`http://www.numpy.org/`)

Software

- Python (`https://www.python.org/`)
- Numpy (`http://www.numpy.org/`)
- Theano (`http://deeplearning.net/software/theano`)

Software

- Python (`https://www.python.org/`)
- Numpy (`http://www.numpy.org/`)
- Theano (`http://deeplearning.net/software/theano`)
- Tensorflow (`https://www.tensorflow.org/`)

Software

- Python (`https://www.python.org/`)
- Numpy (`http://www.numpy.org/`)
- Theano (`http://deeplearning.net/software/theano`)
- Tensorflow (`https://www.tensorflow.org/`)
- Kears (`https://keras.io/`)

Software

- Python (`https://www.python.org/`)
- Numpy (`http://www.numpy.org/`)
- Theano (`http://deeplearning.net/software/theano`)
- Tensorflow (`https://www.tensorflow.org/`)
- Kears (`https://keras.io/`)
- scikit-learn (`http://scikit-learn.org/stable/`)

Neural Network Architectures

- Feed-Forward Network (FFN)

Neural Network Architectures

- Feed-Forward Network (FFN)
- Recurrent Neural Network (RNN)

Neural Network Architectures

- Feed-Forward Network (FFN)
- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)

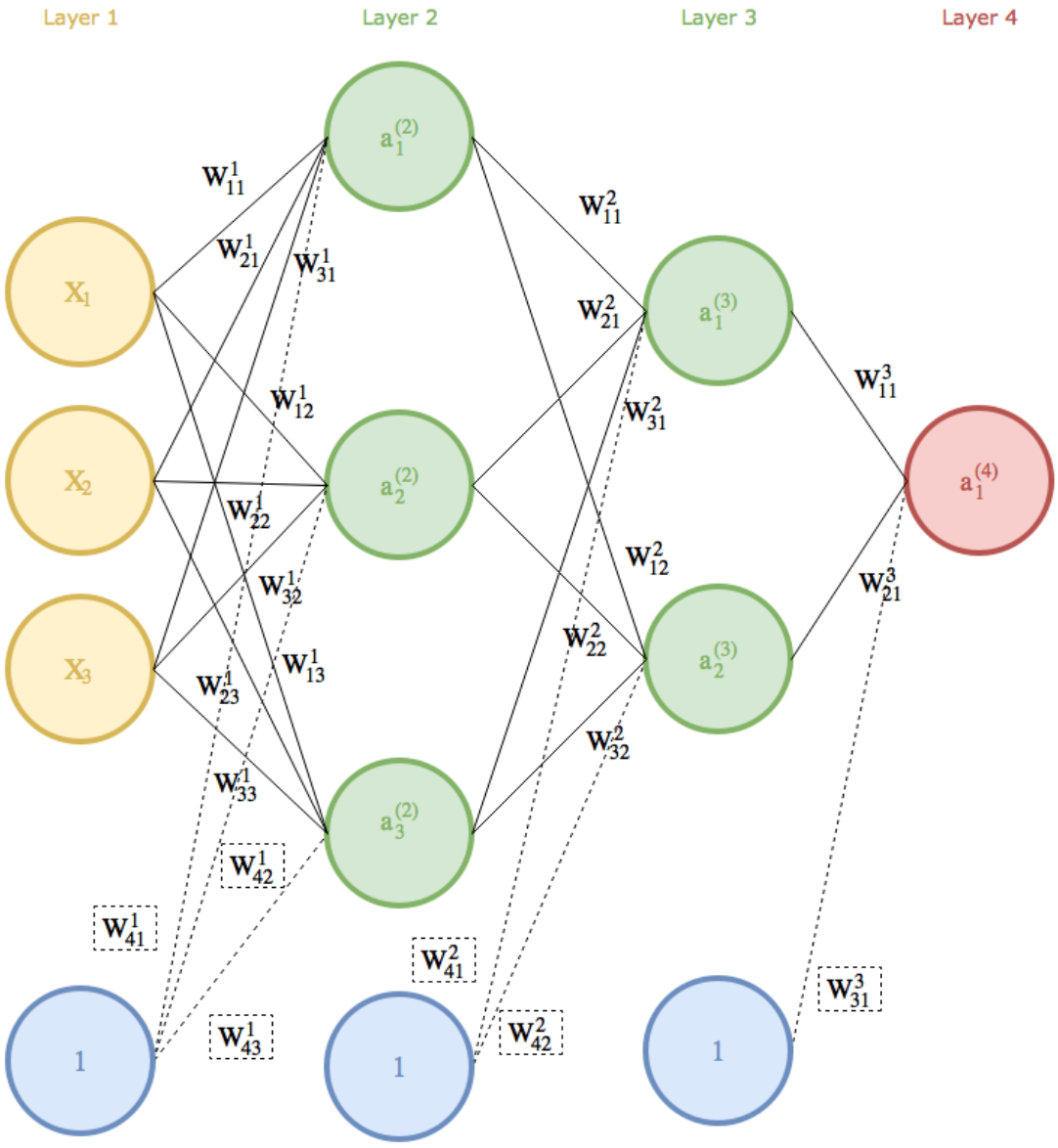
Neural Network Architectures

- Feed-Forward Network (FFN)
- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)
- Bidirectional Recurrent Neural Network (BiLSTM)

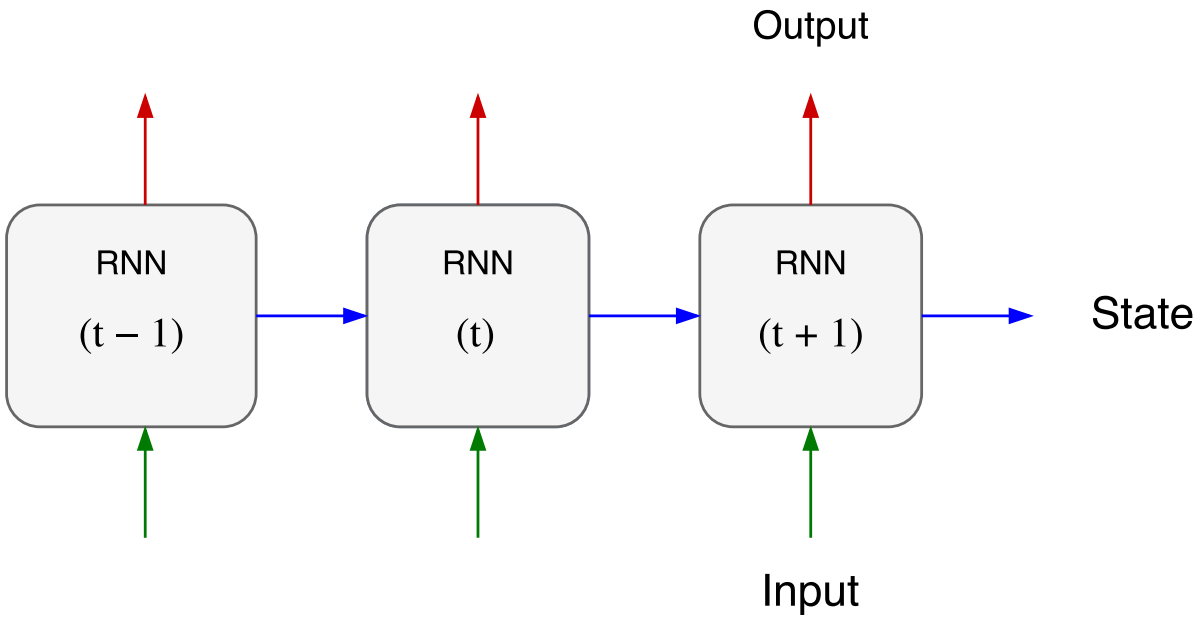
Neural Network Architectures

- Feed-Forward Network (FFN)
- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)
- Bidirectional Recurrent Neural Network (BiLSTM)
- Convolutional Neural Network (CNN)

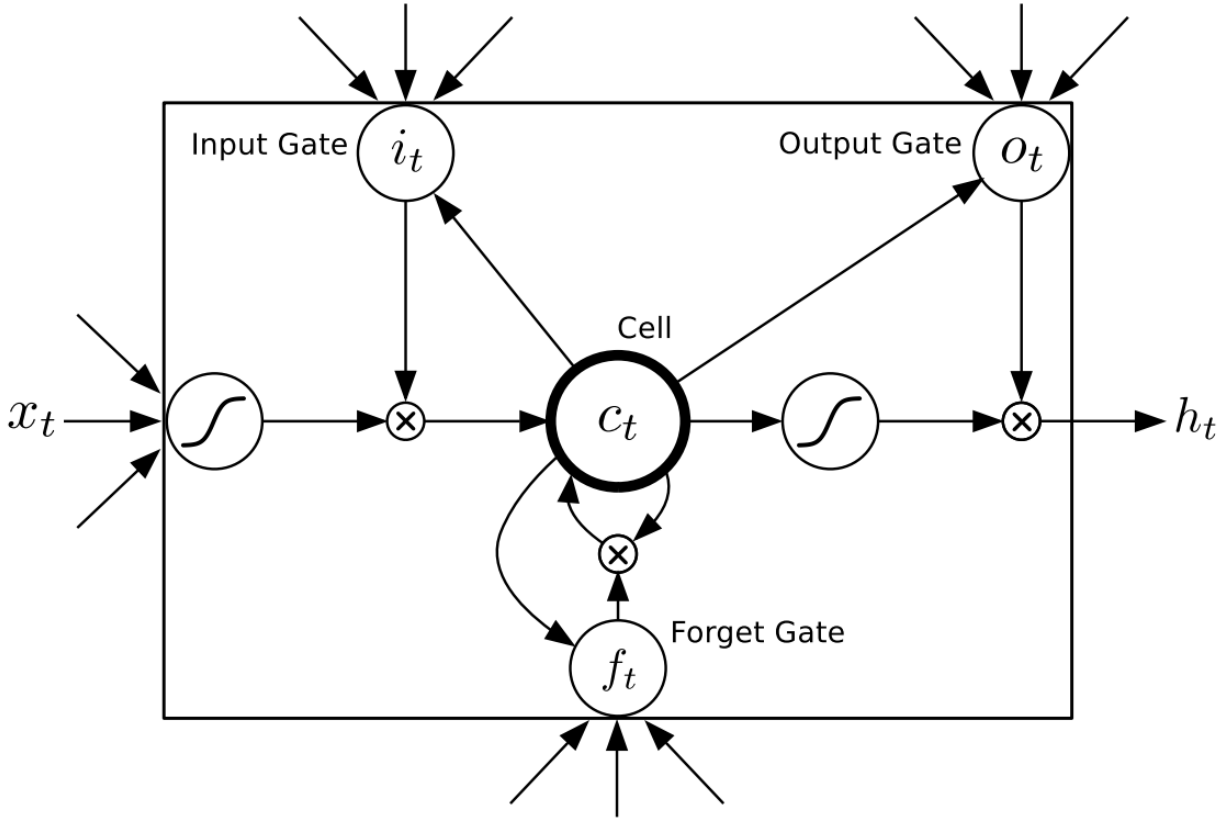
Feed-Forward Network



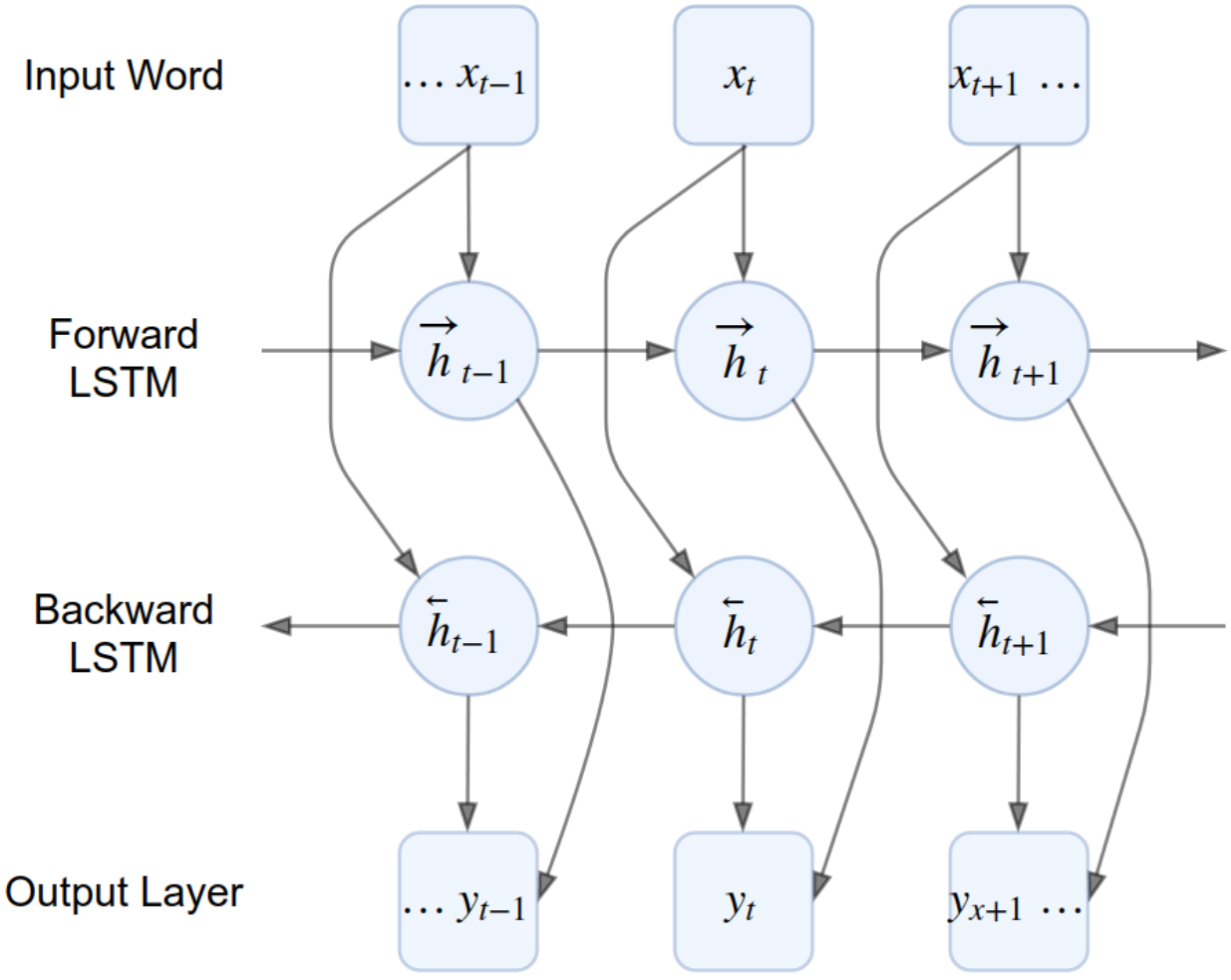
Recurrent Neural Network



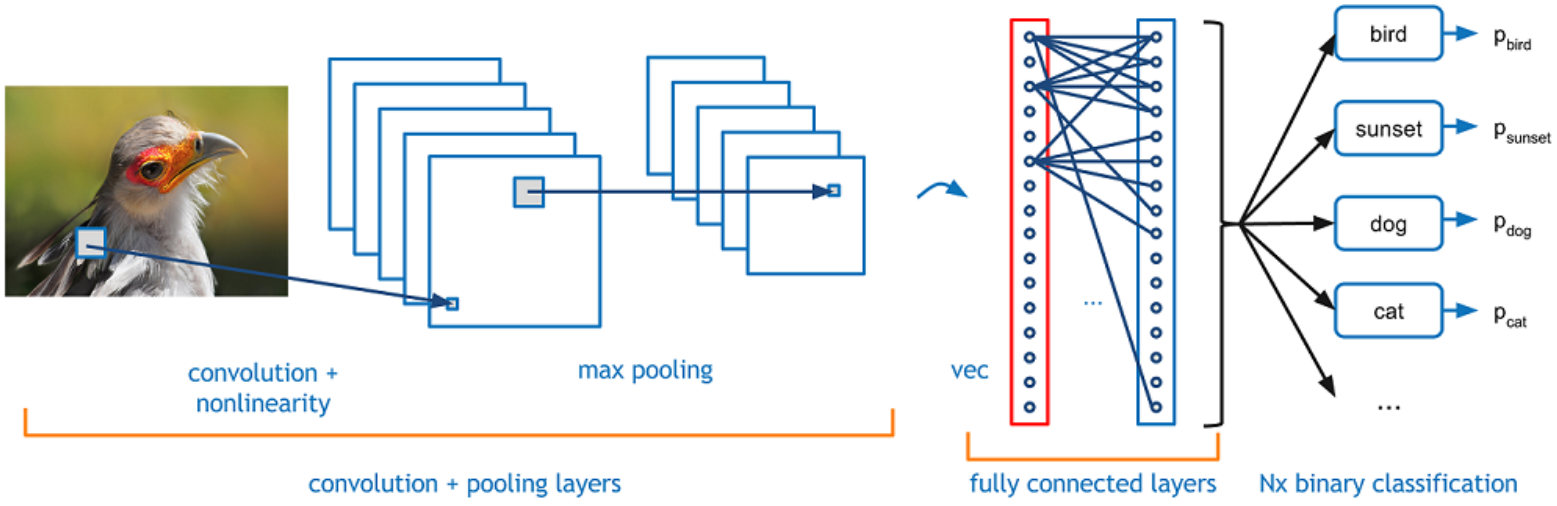
Long Short-Term Memory



Bidirectional Recurrent Neural Network



Convolutional Neural Network



Deep Learning

- A subfield of Machine Learning

Deep Learning

- A subfield of Machine Learning
- Deep Learning (DL) is now emerging as a new research area in ML and NLP

Deep Learning

- A subfield of Machine Learning
- Deep Learning (DL) is now emerging as a new research area in ML and NLP
- Over the last several years, NLP techniques have extensively relied on linear models

Deep Learning

- A subfield of Machine Learning
- Deep Learning (DL) is now emerging as a new research area in ML and NLP
- Over the last several years, NLP techniques have extensively relied on linear models
- Reliance on external resources or carefully hand-crafted features

Deep Learning

- A subfield of Machine Learning
- Deep Learning (DL) is now emerging as a new research area in ML and NLP
- Over the last several years, NLP techniques have extensively relied on linear models
- Reliance on external resources or carefully hand-crafted features
- time-consuming to construct

Deep Learning

- A subfield of Machine Learning
- Deep Learning (DL) is now emerging as a new research area in ML and NLP
- Over the last several years, NLP techniques have extensively relied on linear models
- Reliance on external resources or carefully hand-crafted features
- time-consuming to construct
- over-specified and incomplete

Motivation

- The knowledge learnt from Deep Neural Network (DNN) models is still largely untapped in the context of NLP

Motivation

- The knowledge learnt from Deep Neural Network (DNN) models is still largely untapped in the context of NLP
- DNNs generalize well on unseen data

Motivation

- The knowledge learnt from Deep Neural Network (DNN) models is still largely untapped in the context of NLP
- DNNs generalize well on unseen data
- They are suitable to deal with outlying, missing, unstructured, and noisy data.

Motivation

- The knowledge learnt from Deep Neural Network (DNN) models is still largely untapped in the context of NLP
- DNNs generalize well on unseen data
- They are suitable to deal with outlying, missing, unstructured, and noisy data.
- Very flexible and can be paired with other techniques

Motivation

- The knowledge learnt from Deep Neural Network (DNN) models is still largely untapped in the context of NLP
- DNNs generalize well on unseen data
- They are suitable to deal with outlying, missing, unstructured, and noisy data.
- Very flexible and can be paired with other techniques
- DNNs are also inherently non-linear

Motivation

- The knowledge learnt from Deep Neural Network (DNN) models is still largely untapped in the context of NLP
- DNNs generalize well on unseen data
- They are suitable to deal with outlying, missing, unstructured, and noisy data.
- Very flexible and can be paired with other techniques
- DNNs are also inherently non-linear
- outperform linear models in classification and data analysis tasks

Questions

- Do these representations correspond in any interpretable way to linguistically motivated representations typically used in theoretical linguistics?

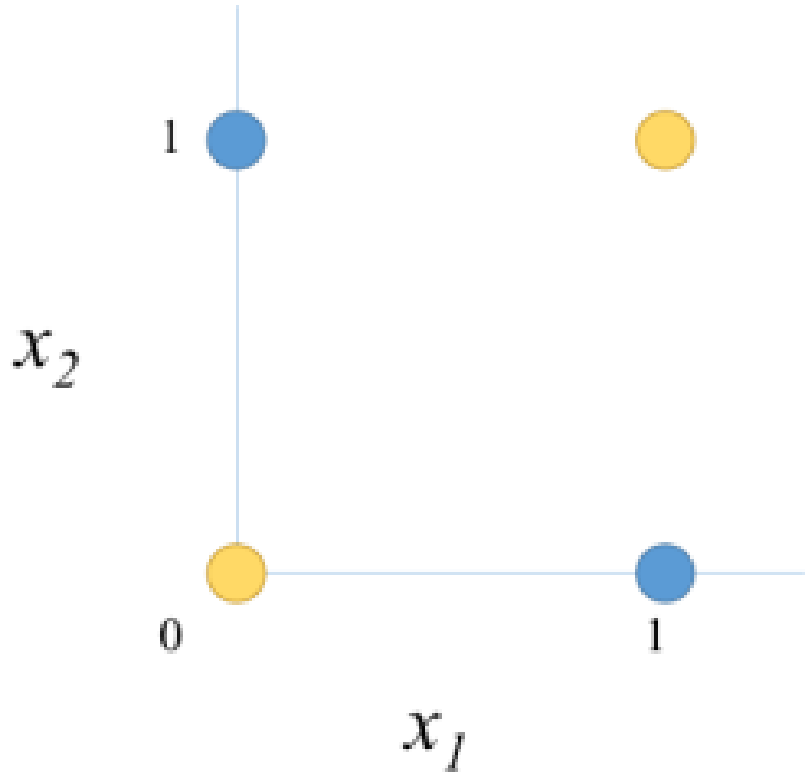
Questions

- Do these representations correspond in any interpretable way to linguistically motivated representations typically used in theoretical linguistics?
- What are the criteria that make one representation better than another?

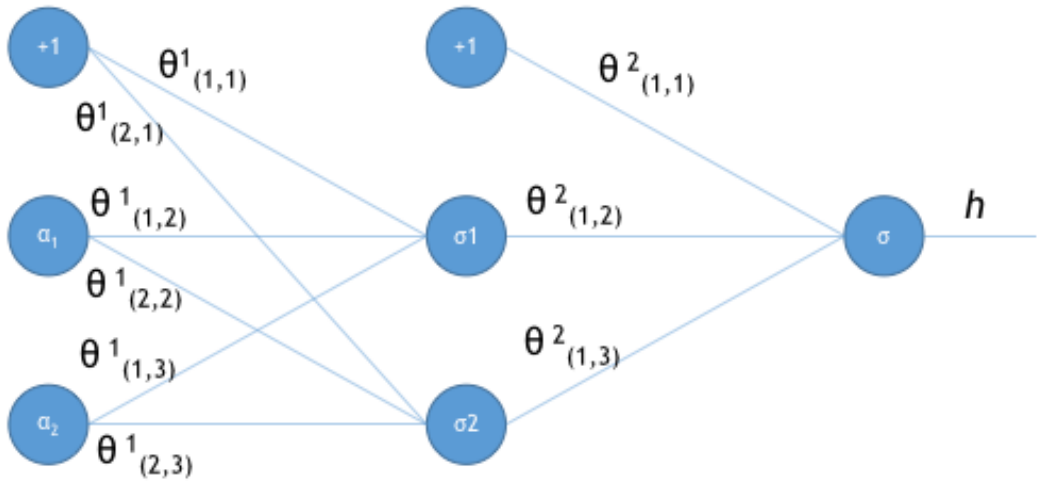
The XOR Problem

Given this input		Produce this output
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

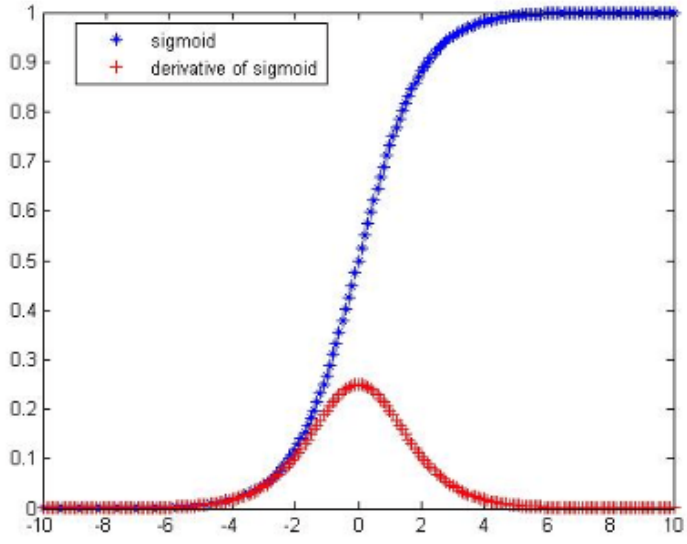
Linear-Separability



MLP


+1
+1
$\theta^1_{(1,1)}$
$\theta^1_{(2,1)}$
$\theta^2_{(1,1)}$
α_1
$\theta^1_{(1,2)}$
$\theta^1_{(2,2)}$
σ1
$\theta^2_{(1,2)}$
σ
h
$\theta^1_{(1,3)}$
$\theta^2_{(1,3)}$
α_2
$\theta^1_{(2,3)}$
σ2

Activation Function

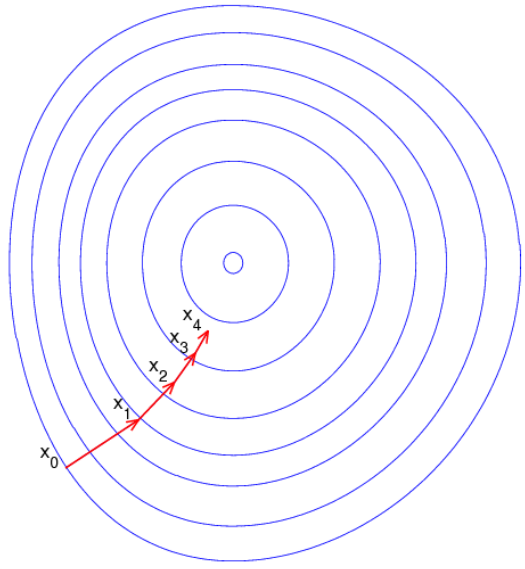


$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

$$\frac{d\sigma(x)}{d(x)} = \sigma(x) \cdot (1 - \sigma(x)) \tag{2}$$

Stochastic Gradient Descent Algorithm

Steepest Descent



Steepest Descent

- Gradient descent is an iterative optimization algorithm for finding the minimum of a function
- Takes steps proportional to the (negative or positive) of the gradient

Stochastic Gradient Descent Algorithm

Pseudocode

```
initialize network weights (often small random values)
do
   forEach training example named ex
      prediction = neural-net-output(network, ex)  // forward pass
      actual = teacher-output(ex)
      compute error (prediction - actual) at the output units
      compute Δw_h for all weights from hidden layer to output layer  // backward pass
      compute Δw_i for all weights from input layer to hidden layer   // backward pass continued
      update network weights // input layer not modified by error estimate
until all examples classified correctly or another stopping criterion satisfied
return the network
```

Figure 1 : Stochastic Gradient Descent Algorithm[1]

[1] Borrowed from Wikipedia

Software installation

PIP

A tool for installing and managing Python packages.

$ sudo apt-get update

$ sudo apt-get -y install python-pip

Another possibility to install pip:

$ curl "https://bootstrap.pypa.io/get-pip.py" -o "get-pip.py"

$ python get-pip.py

$ pip -V

Software installation

Theano

Theano is a Python library that allows you to define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays efficiently.

$ sudo pip install Theano

Software installation

TensorFlow

TensorFlow is an open source software library for numerical computation using data flow graphs.

Installation

$ sudo apt-get install python-dev (for Python 2.7)
$ pip install tensorflow (CPU support (no GPU support))
$ pip install tensorflow-gpu (GPU support)

Software installation

Keras

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

Installation

$ sudo pip install keras

Switching from one backend to another

$HOME/.keras/keras.json