Deep Learning for Text Analytics

Ananth Iyer and Felix Wyss
Applied Research Group
www.genesys.com

{ananth.iyer, felix.wyss}@genesys.com

GENESYS™

Agenda

- Introduction
- Word2vec
- Interaction steering application
- Demo!

Introduction

What is Text Analytics?

- Process of extracting meaning from unstructured text

- Examples:
  - Information retrieval: search engines, chat-bots
  - Context aware advertising
  - Spam detection
  - Sentiment analysis
  - Customer care

What is a neural network?

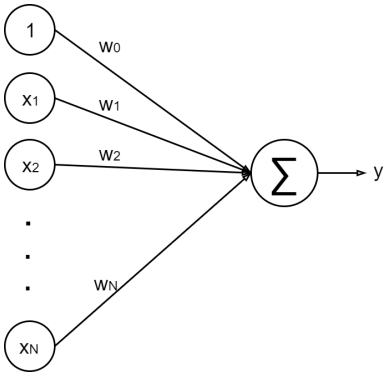
- Neural network is a biologically inspired model which enables a machine to learn from observed data

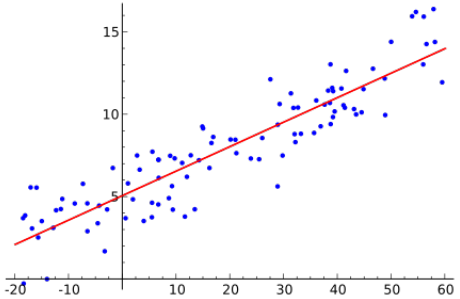
- Neural networks can function as
  - function approximators
  - Probability estimators

Linear/Logistic/Neural Network

- Linear regression

$$y = w_0 + \sum_i w_i x_i$$



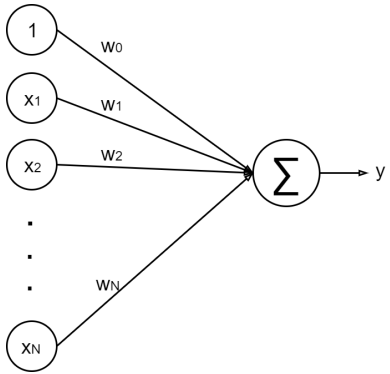


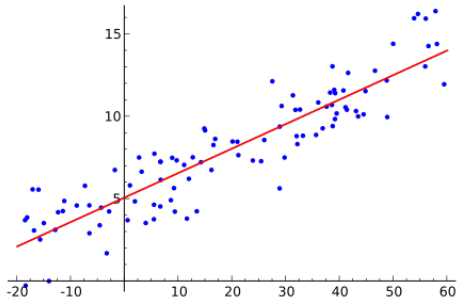
https://en.wikipedia.org/wiki/Linear_regression#/media/File:Linear_regression.svg

Linear/Logistic/Neural Network

- Linear regression

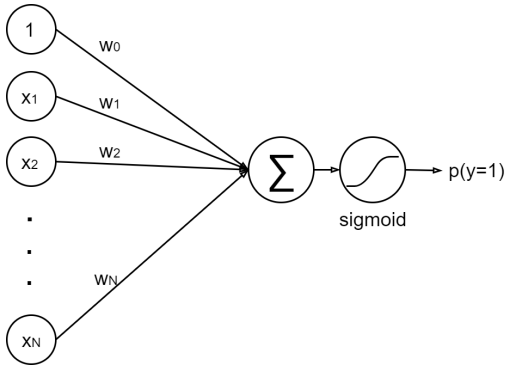
$$y = w_0 + \sum_i w_i x_i$$

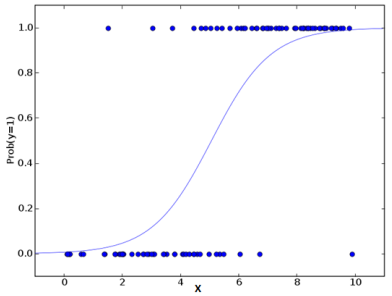




- Logistic regression

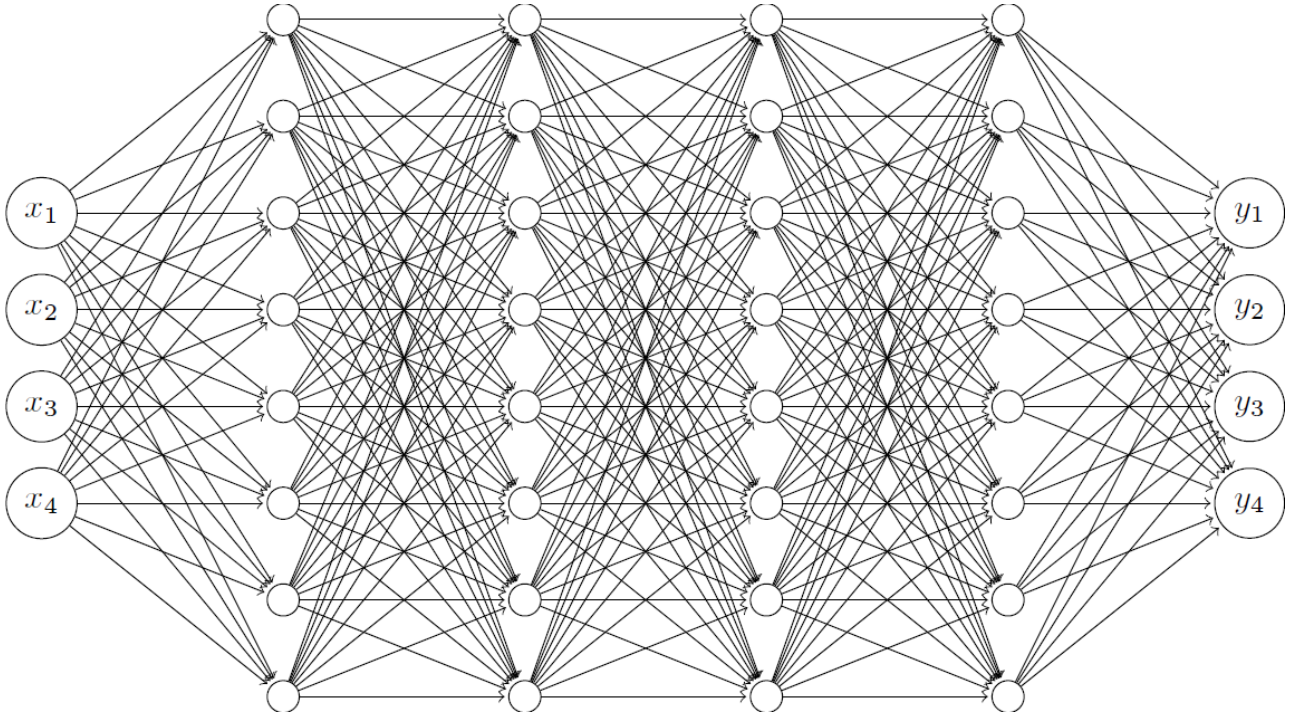
$$P(y = 1) = \frac{1}{1 + e^{-(w_0 + \sum_i w_i x_i)}}$$





https://en.wikipedia.org/wiki/Linear_regression#/media/File:Linear_regression.svg
https://www.analyticsvidhya.com/wp-content/uploads/2015/08/Logistic_Regression.png)

Linear/Logistic/Neural Network



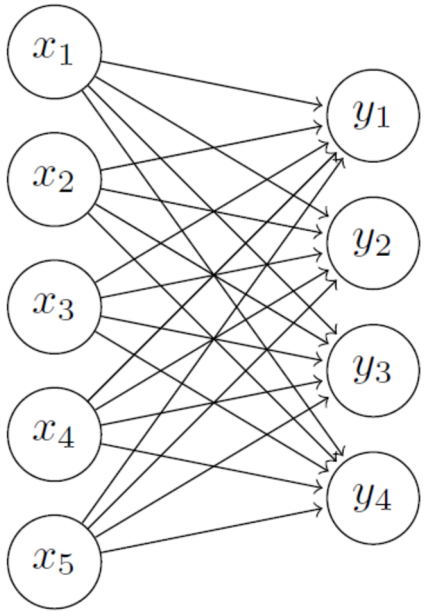
What is deep learning?

- Deep learning is a set of methods to learn data representations (feature learning)

- Neural networks with large number of hidden layers are referred to as deep

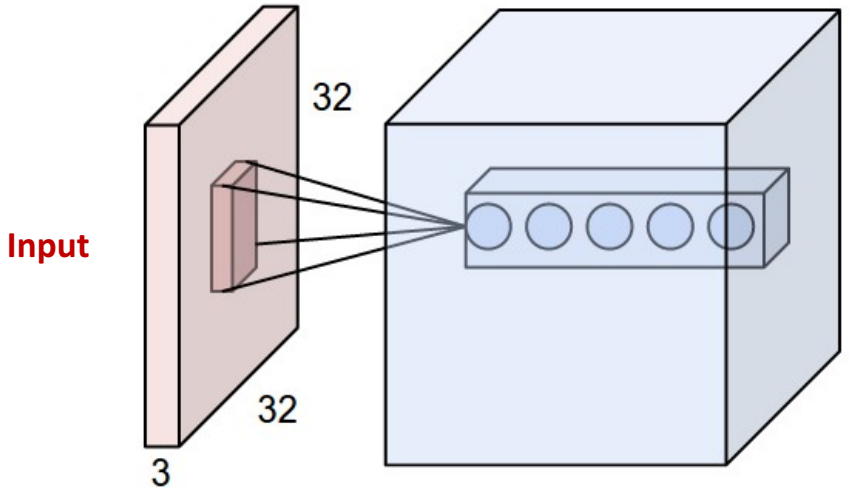
DNN Layer types

- Feed Forward



DNN Layer types

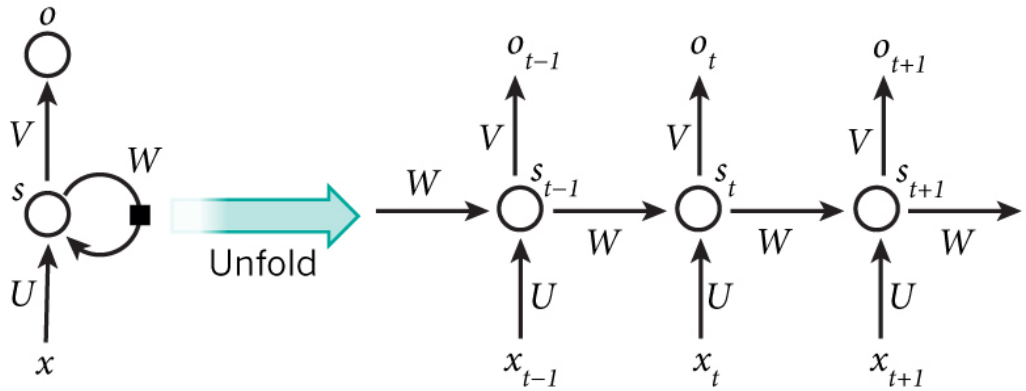
- Feed Forward
- Convolutional



http://cs231n.github.io/assets/cnn/depthcol.jpeg

DNN Layer types

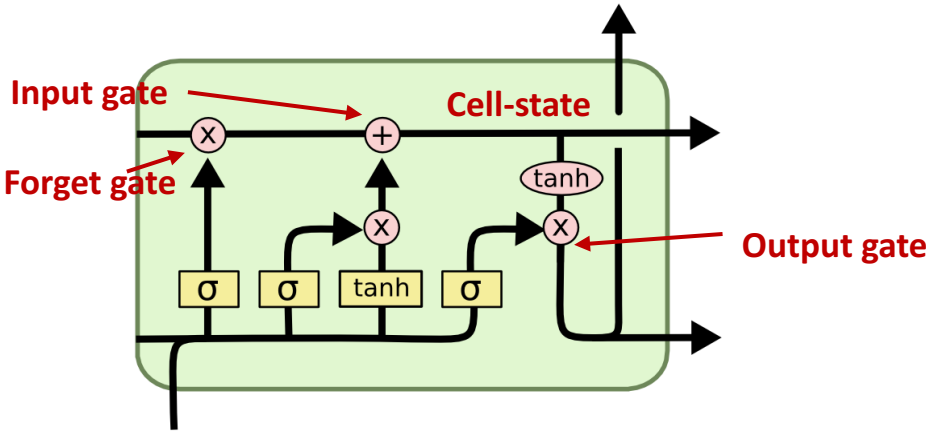
- Feed Forward
- Convolutional
- Recurrent



http://d3kbpzbmcynnmx.cloudfront.net/wp-content/uploads/2015/09/rnn.jpg

DNN Layer types

- Feed Forward
- Convolutional
- Recurrent
- Long short-term memory



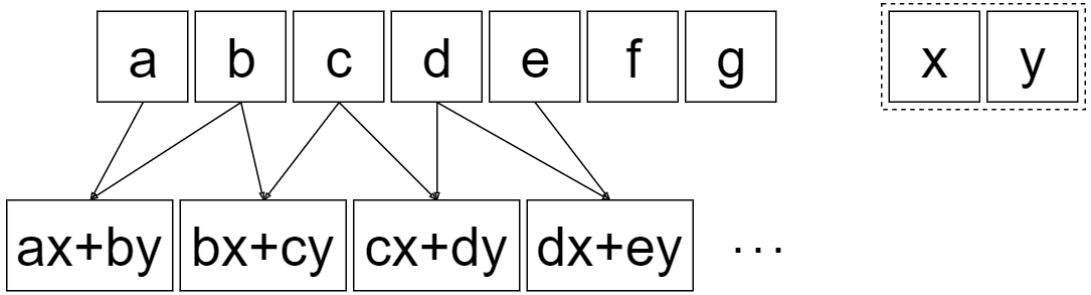
http://colah.github.io/images/post-covers/lstm.png

Convolutional

- Scales up to very large inputs
  - Sparse connections
  - Parameter sharing

Convolutional

- Scales up to very large inputs
  - Sparse connections
  - Parameter sharing



word2Vec

Semantic vector space representation of words

GENESYS

One-hot encoding of words

- Common approach to encode a word as vector is:

```
[0 0 0 0 0 ... 0 0 0 1 0 0]
```

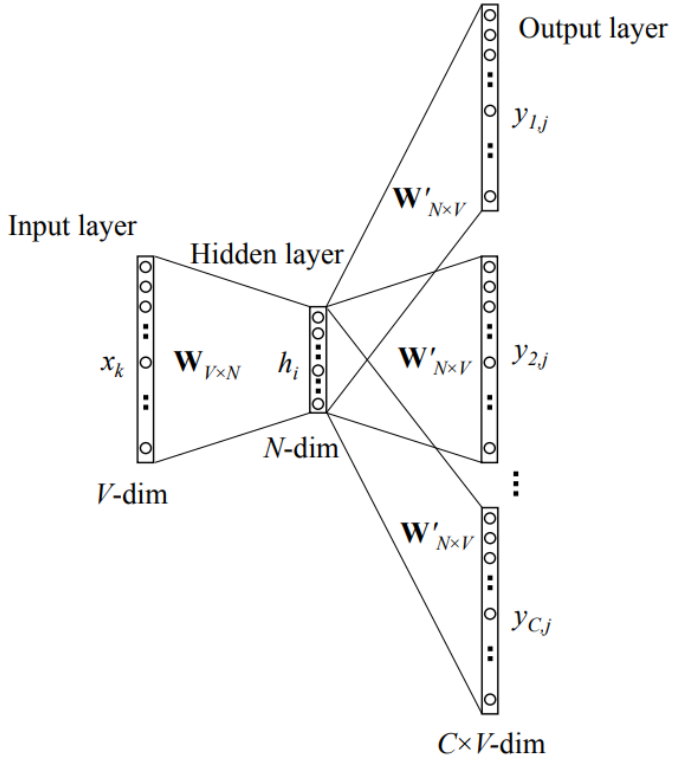
- What is the problem?
  - Vectors are very sparse
  - Vector dimensions can be really large $\propto$ vocabulary size

Word embedding

- Reduces dimensionality
- Learned using a simple (2-layer) neural network
- Embedded vectors maintain semantic relationship

- Example:

  `E('king')-E('man') ≈ E('queen')-E('woman')`



https://arxiv.org/pdf/1411.2738.pdf

Text pre-processing

- Word tokenization

```
>>> nltk.tokenize.word_tokenize('I like Python! I really do.')
['I', 'like', 'Python', '!', 'I', 'really', 'do', '.']
```

- Stemming

```
>>> nltk.stem.porter.PorterStemmer().stem('ponies')
'poni'
```

- Lemmatization

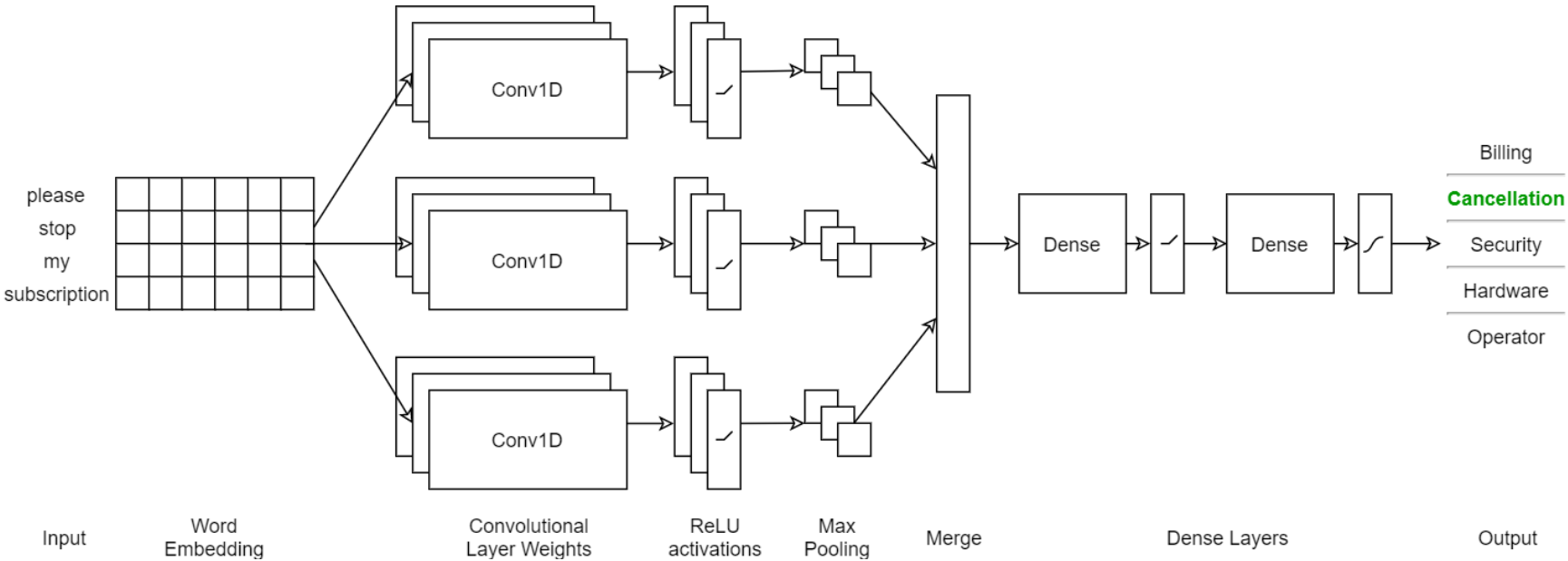
```
>>> nltk.stem.WordNetLemmatizer().lemmatize('ponies')
'pony'
```

- Stop word removal

Interaction steering application

GENESYS

Interaction steering application



Demo

GENESYS

Thank You

Visit www.genesys.com or call +1.855.821.0932 for more information

GENESYS