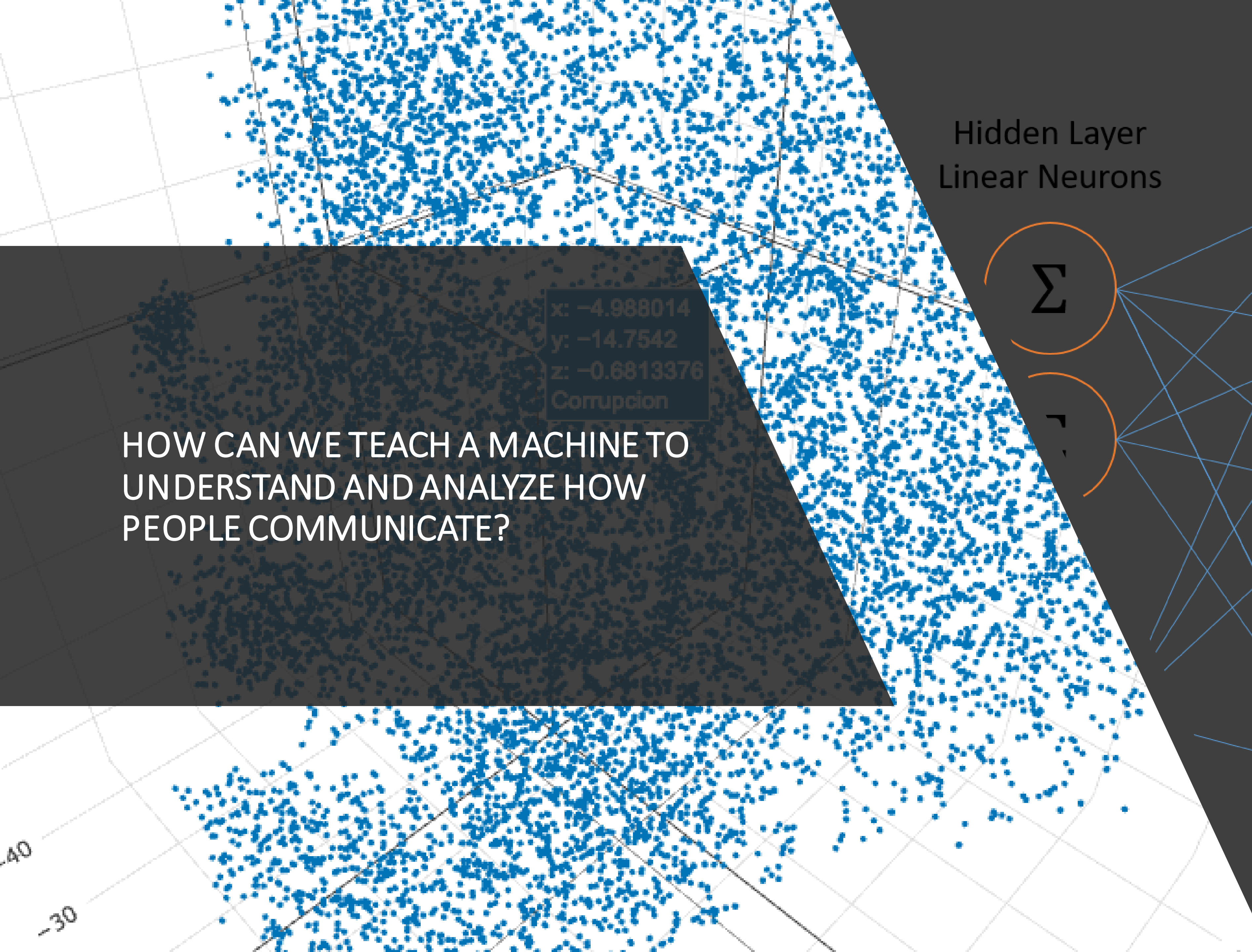
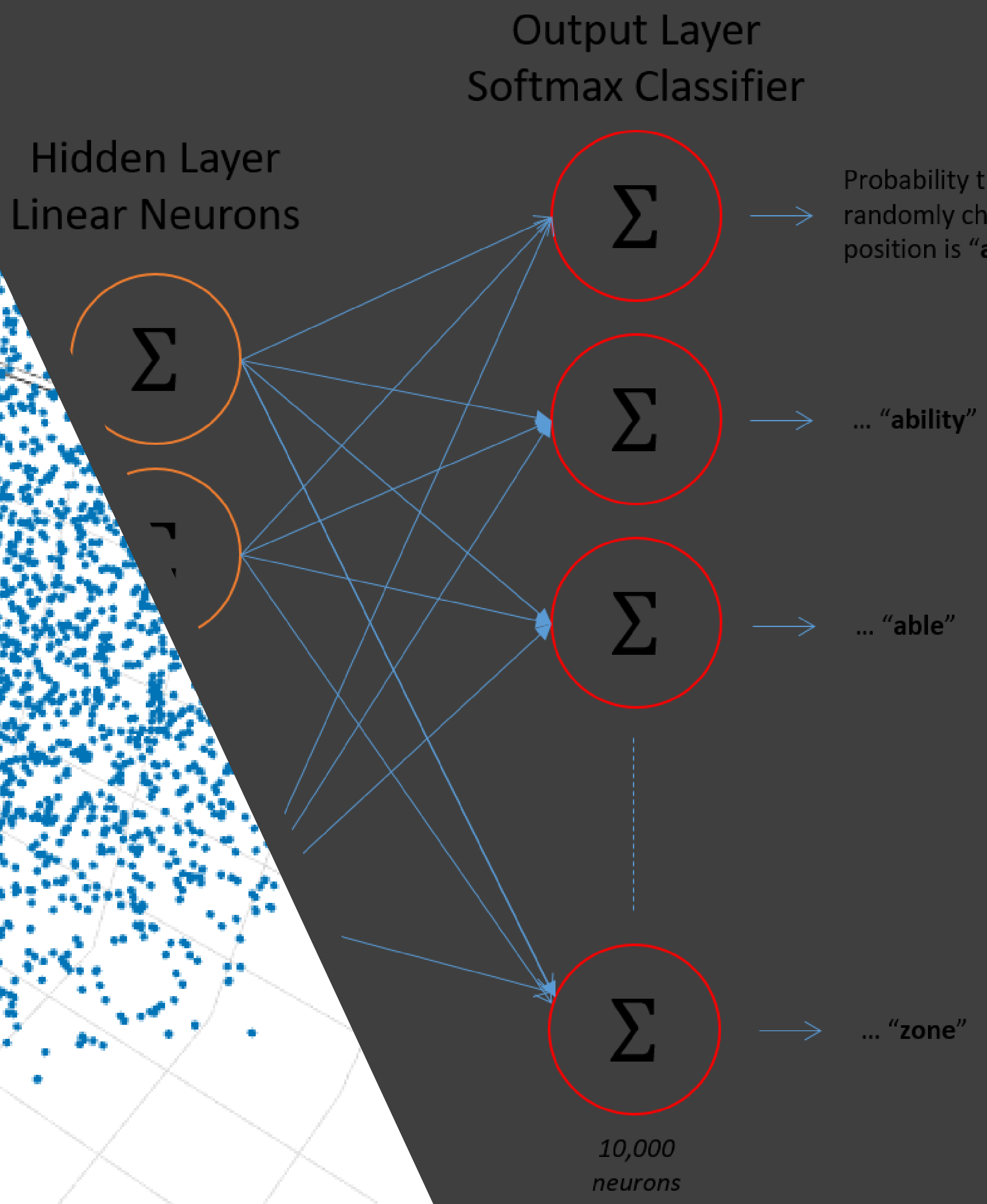
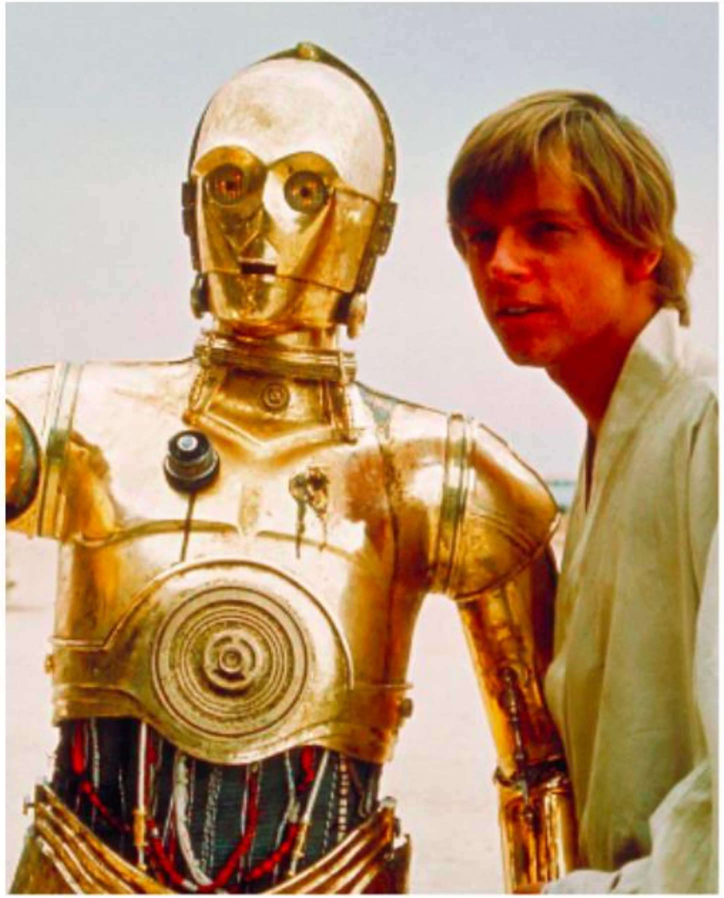
HOW CAN WE TEACH A MACHINE TO UNDERSTAND AND ANALYZE HOW PEOPLE COMMUNICATE?





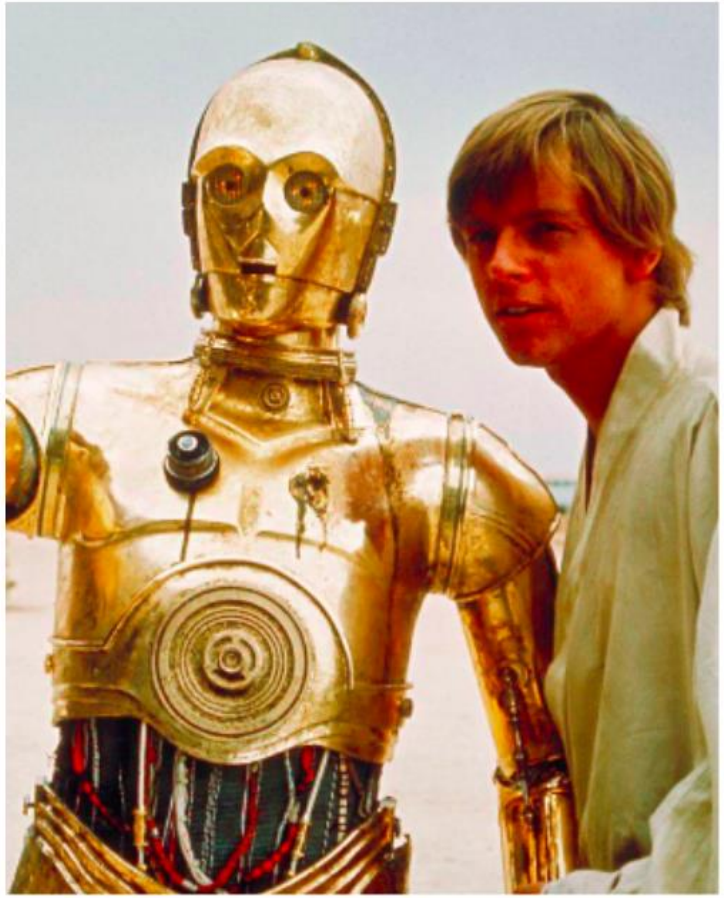
Talking to machines

Expectation



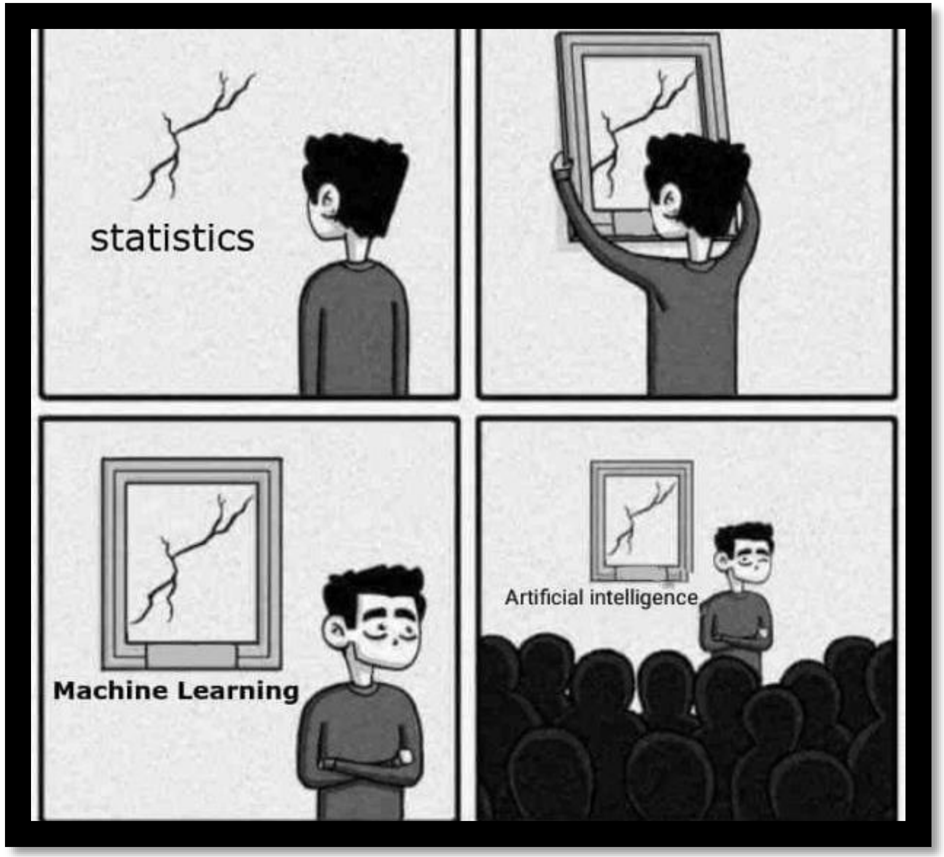
Talking to machines

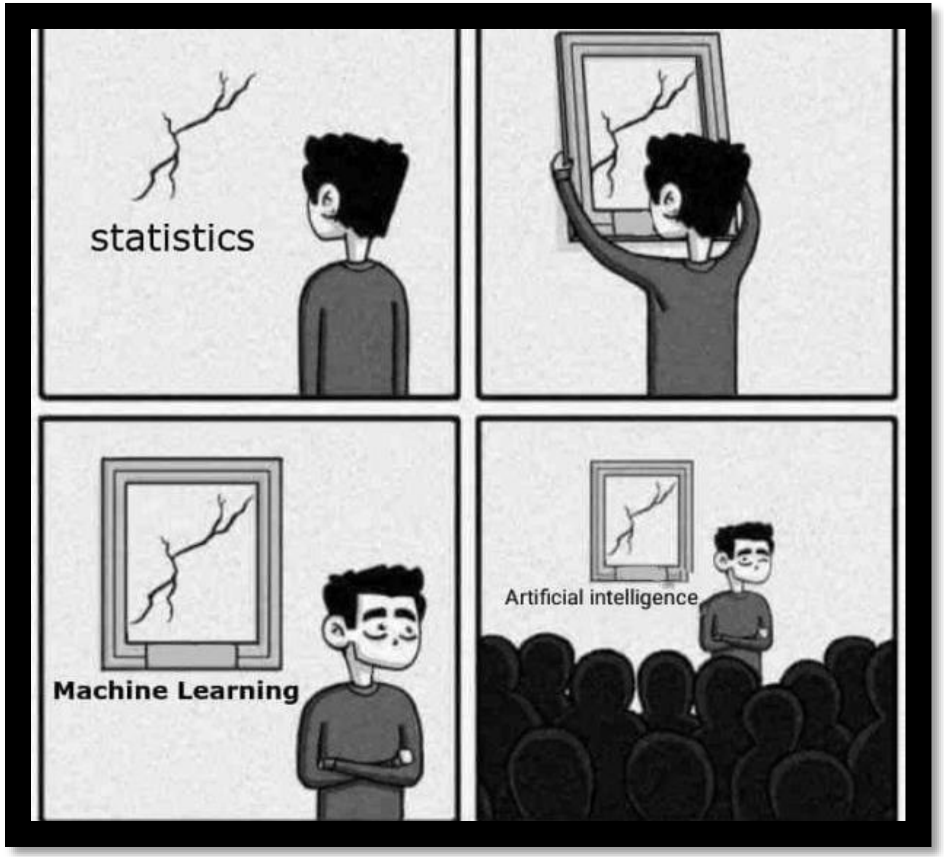
Expectation



Reality

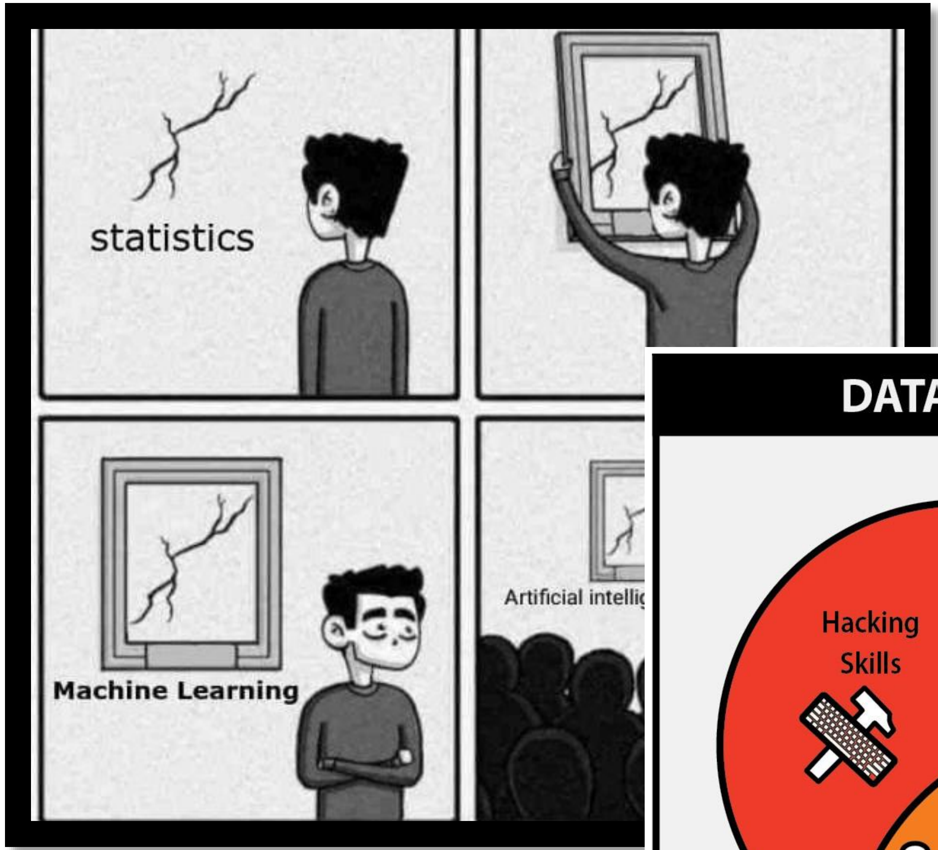
```
X = np.array([ [0,0,1],[0,1,1],[1,0,1],[1,1,1] ])
y = np.array([[0,1,1,0]]).T
syn0 = 2*np.random.random((3,4)) - 1
syn1 = 2*np.random.random((4,1)) - 1
for j in xrange(60000):
    l1 = 1/(1+np.exp(-(np.dot(X,syn0))))
    l2 = 1/(1+np.exp(-(np.dot(l1,syn1))))
    l2_delta = (y - l2)*(l2*(1-l2))
    l1_delta = l2_delta.dot(syn1.T) * (l1 * (1-l1))
    syn1 += l1.T.dot(l2_delta)
    syn0 += X.T.dot(l1_delta)
```

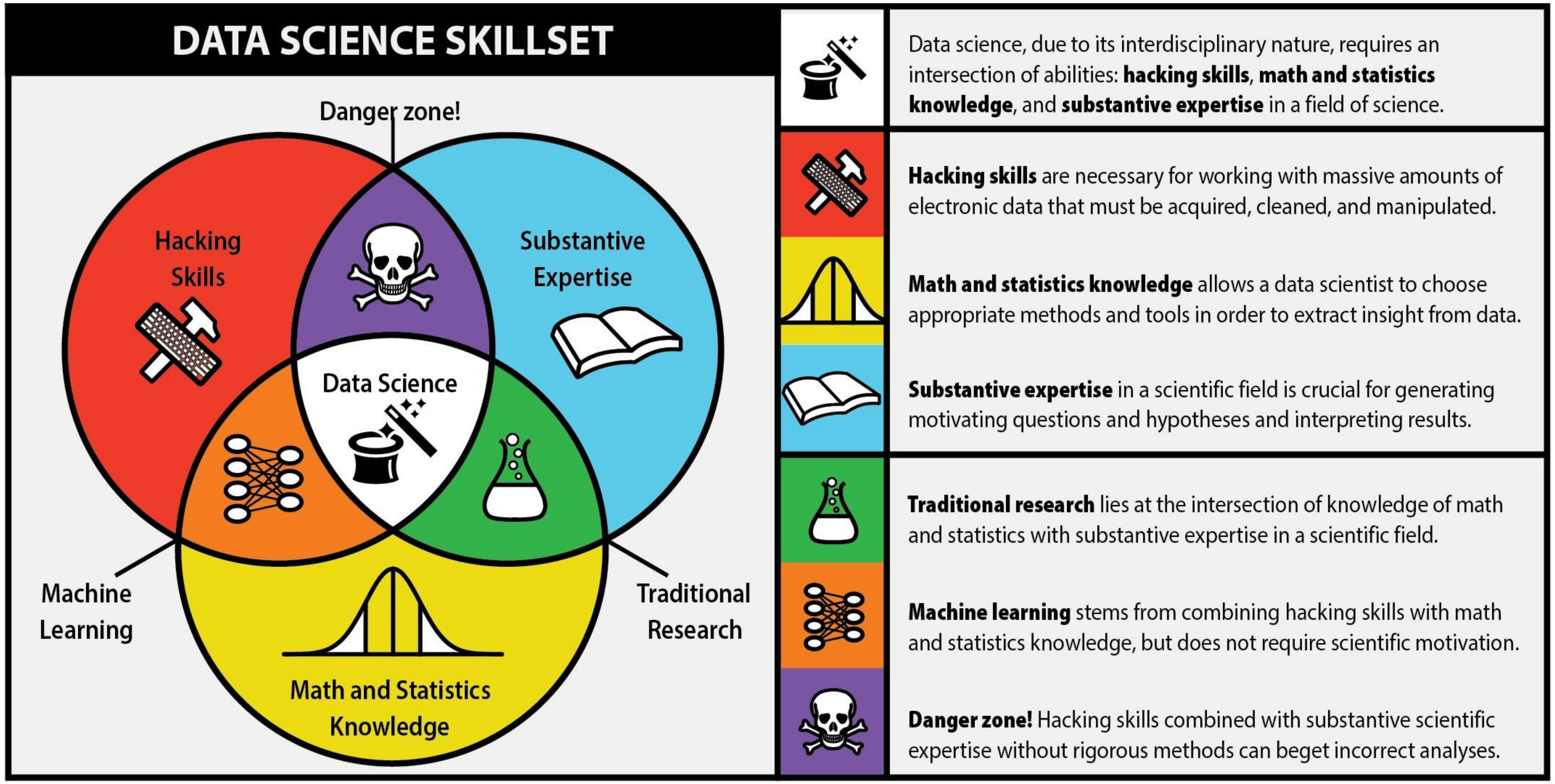

statistics
Machine Learning
Artificial intelligence



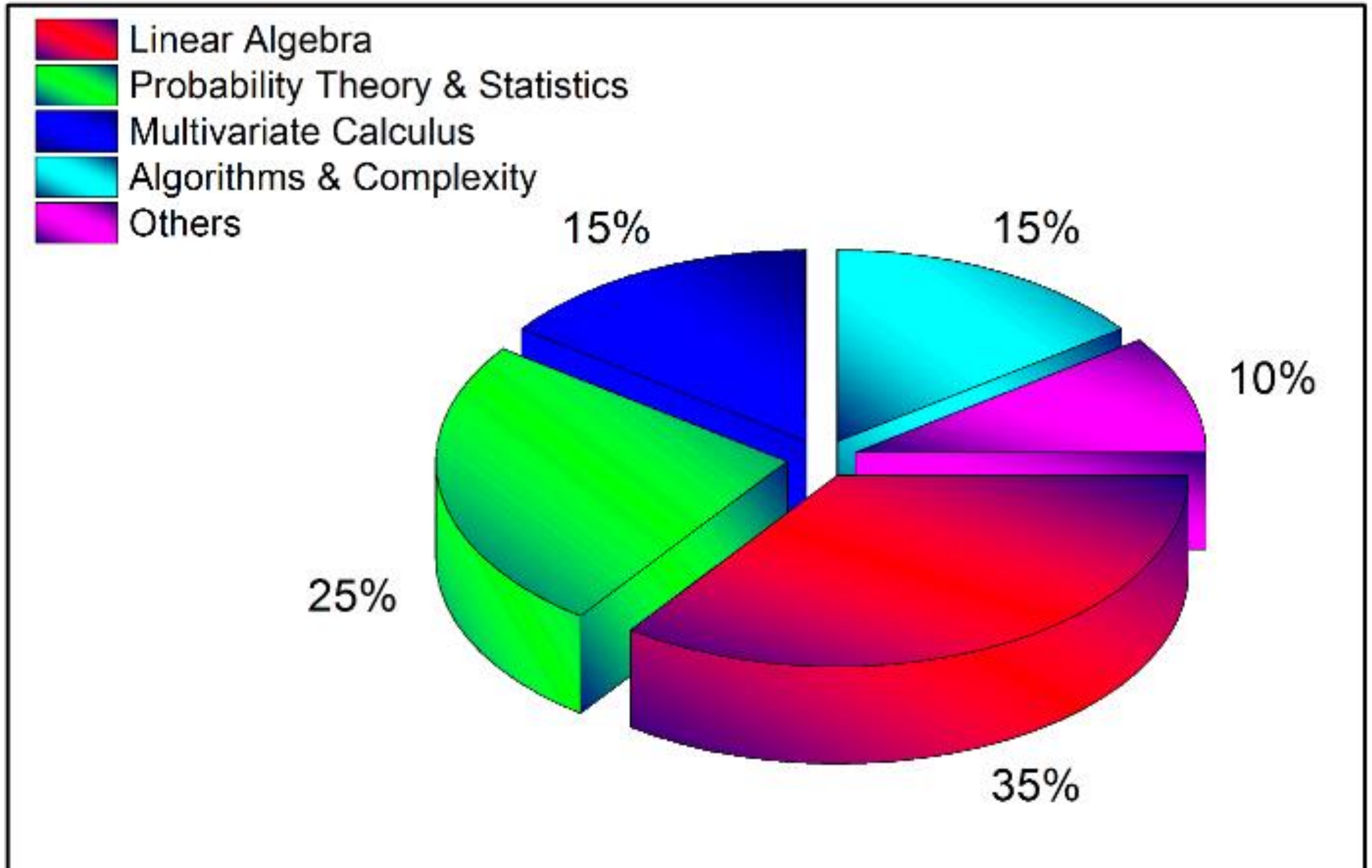
ML is much more than that !

ML is much more than that !





ML is much more than



Importance of Maths Topics Needed For Machine Learning

Data science, due to its interdisciplinary nature, requires an intersection of abilities: **hacking skills**, **math and statistics knowledge**, and **substantive expertise** in a field of science.

Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.

Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insight from data.

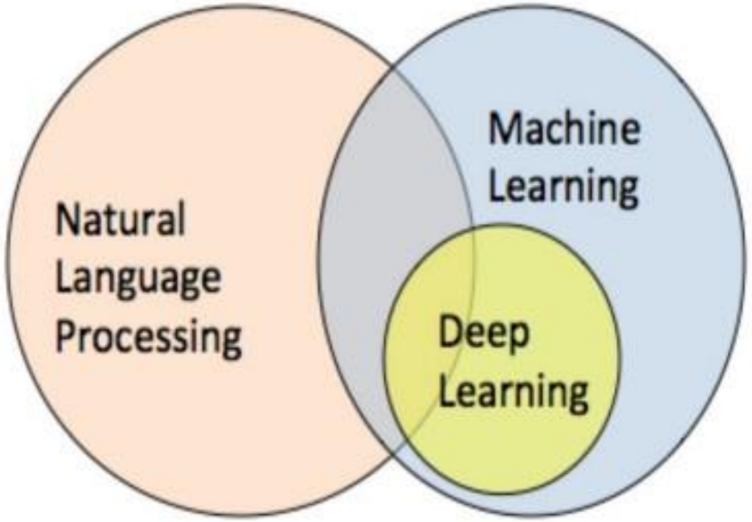
Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.

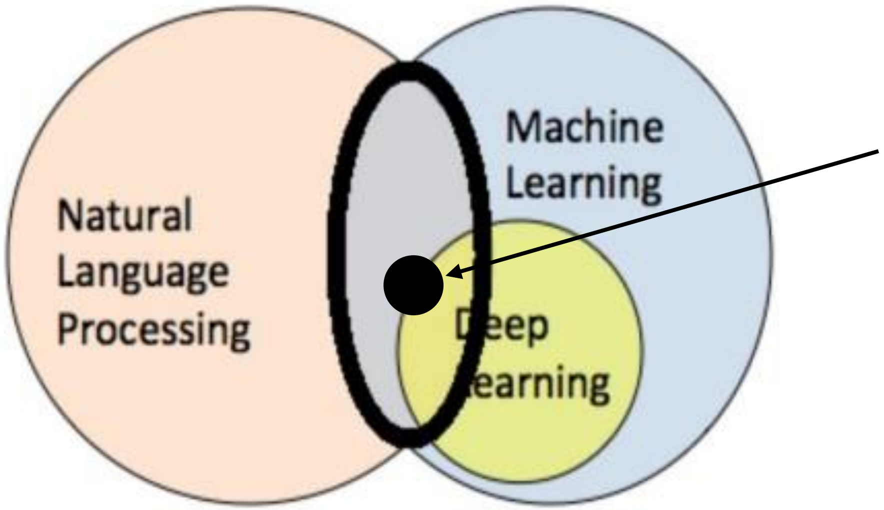
Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.

Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.

Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Where are we in this lecture ?





Focus area

What is a word embedding ?

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

Word embedding

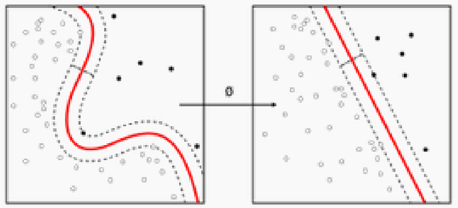
From Wikipedia, the free encyclopedia

Word embedding is the collective name for a set of language modeling and feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers. Conceptually it involves a mathematical embedding from a space with one dimension per word to a continuous vector space with a much lower dimension.

Methods to generate this mapping include neural networks,[1] dimensionality reduction on the word co-occurrence matrix,[2][3][4] probabilistic models,[5] explainable knowledge base method,[6] and explicit representation in terms of the context in which words appear.[7]

Word and phrase embeddings, when used as the underlying input representation, have been shown to boost the performance in NLP tasks such as syntactic parsing[8] and sentiment analysis.[9]

Machine learning and data mining



Problems [show]

Supervised learning [show]
(**classification** • **regression**)

Encoding Human written Language

- **One-hot encoding** is a highly-dimensional encoding (simplest technique but not recommended)
- Similar words are orthogonal (not possible to define <u>word similarity</u>)

$$\begin{array}{ll} \text{size} & [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0]^T \\ \text{capacity} & [0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0] = 0 \end{array}$$

Encoding Human written Language

A solution via distributional similarity-based representations

You can get a lot of value by representing a word by means of its neighbors

"You shall know a word by the company it keeps"

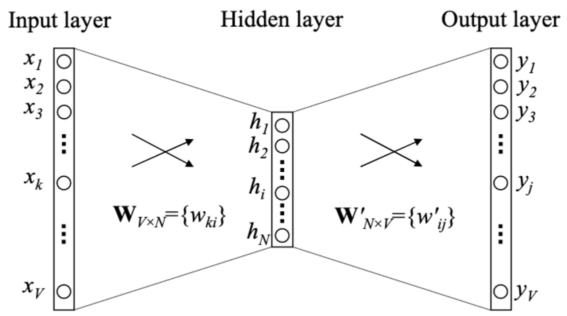
(J. R. Firth 1957: 11)

One of the most successful ideas of modern NLP

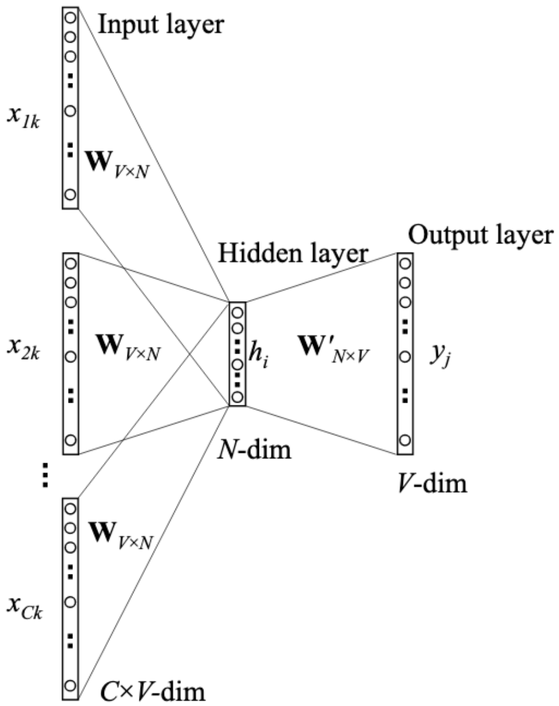
government debt problems turning into **banking** crises as has happened in
saying that Europe needs unified **banking** regulation to replace the hodgepodge

↖ These words will represent *banking* ↗

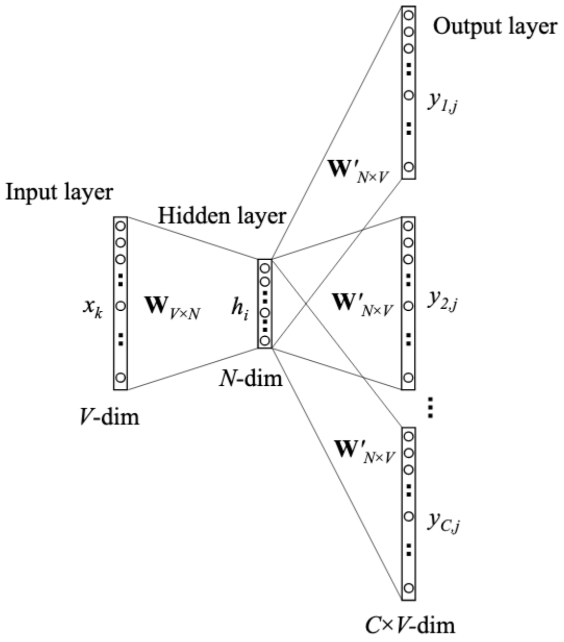
Word2Vec family ….



Bigram CBOW

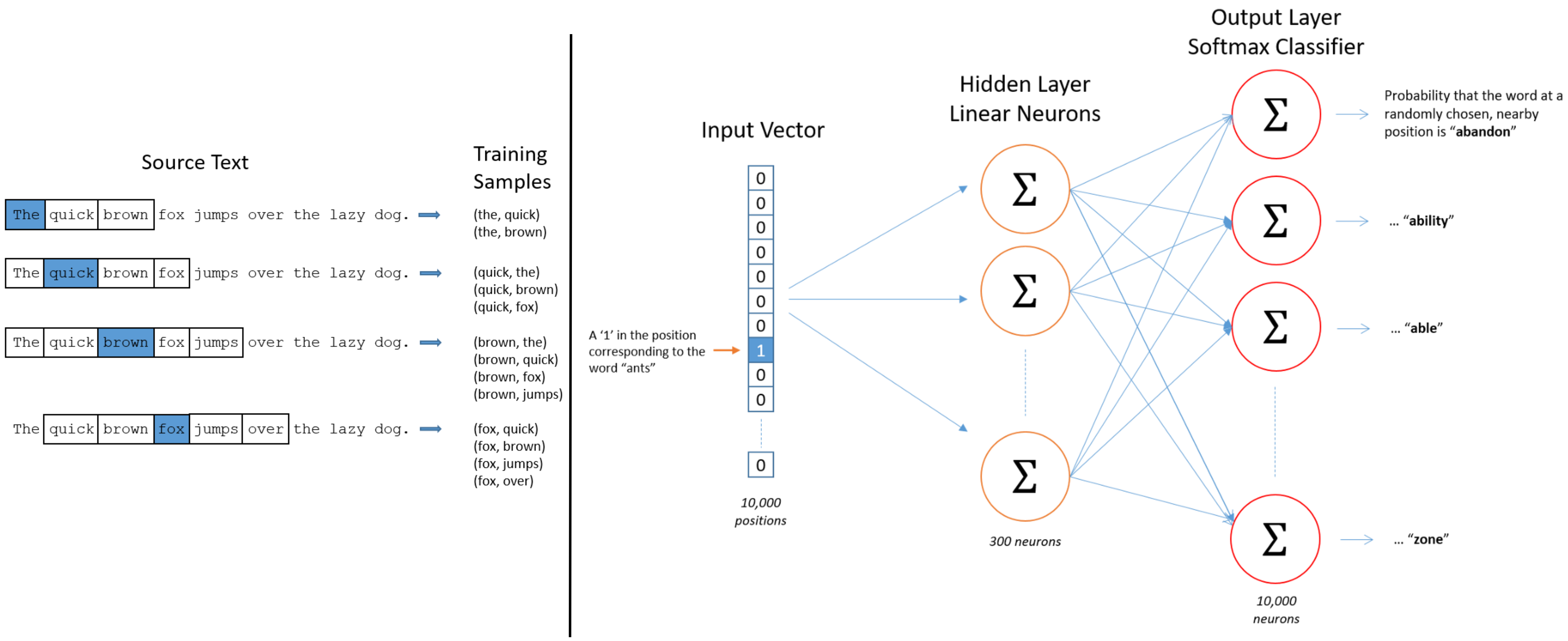


CBOW



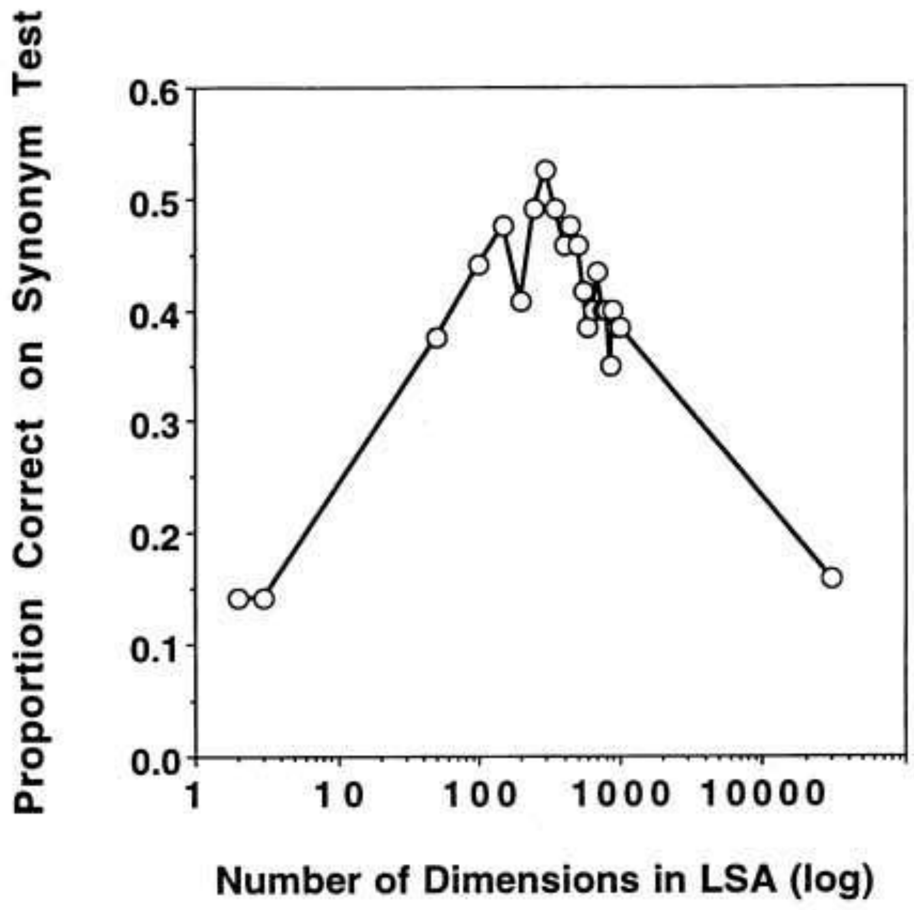
Skip-Gram

A friendly view ...



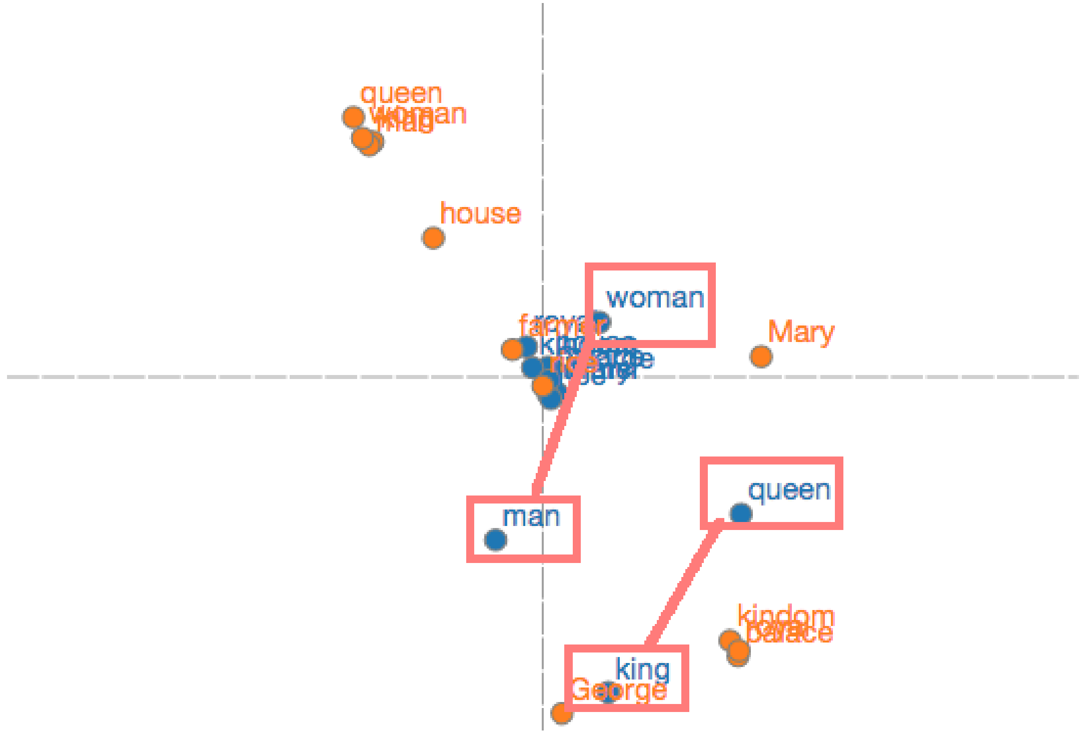
Choosing the dimension of the representation space

- Ex: A striking finding in empirical work on word embeddings is that there is a sweet spot for the dimensionality of word vectors: neither too small, nor too large ([Latent Semantic Analysis paper]()).



Why do semantic relations represent directions ?

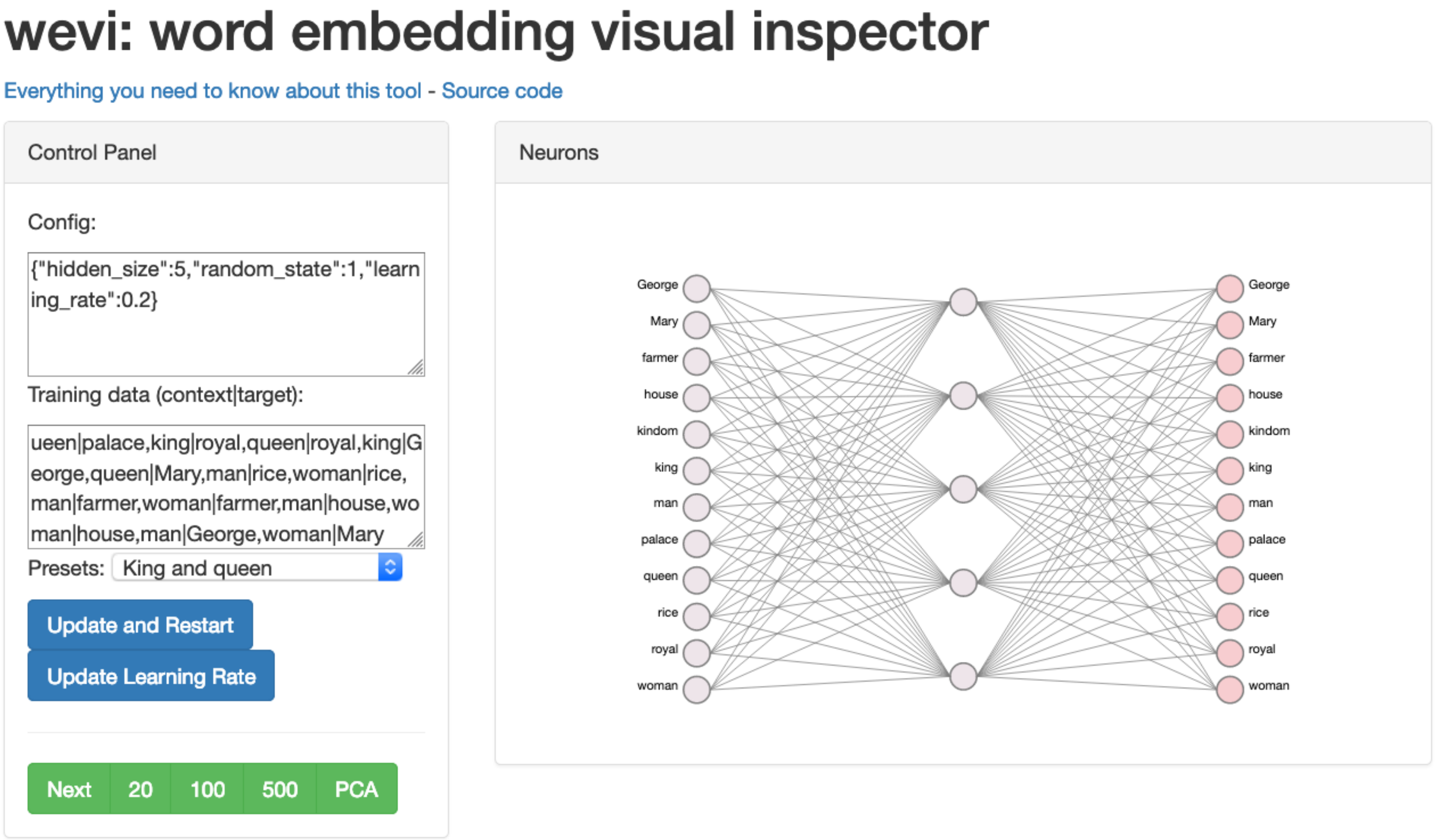
- Striking discovery in the word2vec paper: word analogy tasks can be solved by simple linear algebra.
- semantic relations correspond to directions in representation vector space.



Word2Vec Visual Inspector

[Web inspector]

[Documentation]



References (easy to read!)

- The Mathematics of Machine Learning - Wale Akinfaderin (Medium)
- No, Machine Learning is not just glorified Statistics - Joe Davison (Medium)
- The Statistical Foundation For Artificial Intelligence - John C. Dogan (Medium)
- Understanding Natural Language Understanding – Bill MacCartney (Stanford Lecture)
- Representations for Language: From Word Embeddings to Sentence Meanings - Christopher Manning (Stanford Lecture)
- Chapter 9 : Natural Language Processing - Madhu Sanjeevi (Medium)
- Word2Vec (skip-gram model): PART 1 – Intuition (Medium)

References (more Technical !)

- Word Embeddings: Explaining their properties - Sanjeev Arora
- Semantic Word Embeddings - Sanjeev Arora
- word2vec Parameter Learning Explained - Xin Rong (arxiv)
- Stanford Deep Learning Notes