Word embeddings

One hot encoding

```
aback          [1,0,0,0,0,0,...,0,0,0,0,0]
abacterial     [0,1,0,0,0,0,...,0,0,0,0,0]
abandon        [0,0,1,0,0,0,...,0,0,0,0,0]
abbot          [0,0,0,1,0,0,...,0,0,0,0,0]

...

zoologist      [0,0,0,0,0,0,...,0,1,0,0,0]
zoom           [0,0,0,0,0,0,...,0,0,1,0,0]
zucchini       [0,0,0,0,0,0,...,0,0,0,1,0]
zulu           [0,0,0,0,0,0,...,0,0,0,0,1]
```

Bag of words

“The abacterial zoologist was taken aback by the abandoned zucchini.”

```
aback          [1,0,0,0,0,0,...,0,0,0,0,0]
abacterial     [0,1,0,0,0,0,...,0,0,0,0,0]
abandon        [0,0,1,0,0,0,...,0,0,0,0,0]
...
zoologist      [0,0,0,0,0,0,...,0,1,0,0,0]
zucchini       [0,0,0,0,0,0,...,0,0,0,1,0]

Sum:           [1,1,1,0,0,0,...,0,1,0,1,0]
```

Problems

- “*The man is eating a pizza.*” == “*The pizza is eating a man.*”

- “*car*” & “*automobile*” are as far apart as “*linguist*” and “table”.

- Sparse but HUGE vectors (enumerate every word in a language)

Distributional hypothesis

"*a word is characterized by the company it keeps*"

- J. R. Firth (1950)

"*Words that occur in the same contexts tend to have similar meanings*"

- Z. S. Harris (1954)

Language is a hard problem

"*Exxon Valdez was an **oil** tanker that gained notoriety...*"

"*Wells are drilled into **oil** reservoirs to extract the ...*"

"*Supplies of **oil** and gas...*"

Words with similar meanings

"*But the ultimate **goal** is to remove...*"

"*We hope to meet our **target** for this...*"

Words with multiple meanings

"*Sports like archery and **target** shooting...*"

word2vec

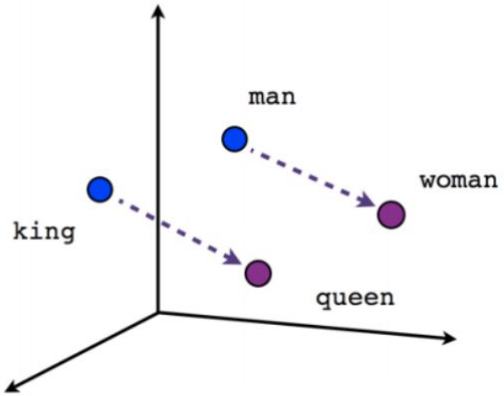
Mikolov T, Le QV, Sutskever I. Exploiting similarities among languages for machine translation. (2013)

- Numeric vector representation of words

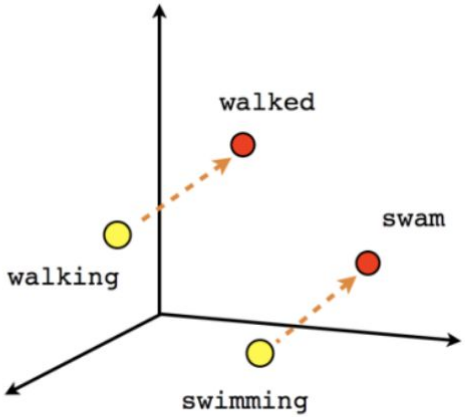
word2vec

- Numeric vector representation of words
- That encapsulate (some) of their meaning

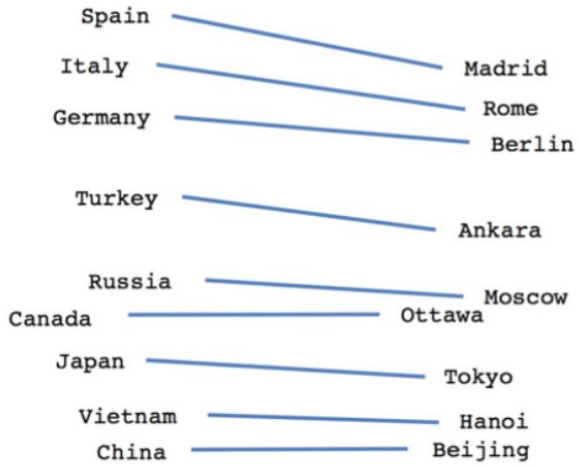
https://www.ruder.io/a-review-of-the-recent-history-of-nlp/



Male-Female



Verb tense



Country-Capital

word2vec Embeddings

```
king =  [0.126,  0.029,  0.008, … , -0.085,  0.091,  0.252]
man =   [0.326,  0.130,  0.034, … , -0.302, -0.080,  0.020]
woman = [0.243, -0.077, -0.103, … , -0.083,  0.065, -0.029]


point = king - man + woman
point ~ queen
queen = [0.005, -0.143, -0.069, … , -0.046,  0.163,  0.154]
```

Training word2vec