Taking uncertainty seriously

A Bayesian approach to word embedding bias estimation

Alicja Dobrzeniecka & Rafal Urbaniak
(LoPSE research group, University of Gdansk)

Boston, April Fools' Day

Word2vec

Question

How to sensibly represent words with numbers?

Word2vec

Question

How to sensibly represent words with numbers?

One-hot encoding

Well, you could use 30k binary vectors with a slot for each lexical unit...

Word2vec

Question

How to sensibly represent words with numbers?

One-hot encoding

Well, you could use 30k binary vectors with a slot for each lexical unit. . .

. . . . . . but this would be inefficient and wouldn't capture any relations between words.

Word2vec

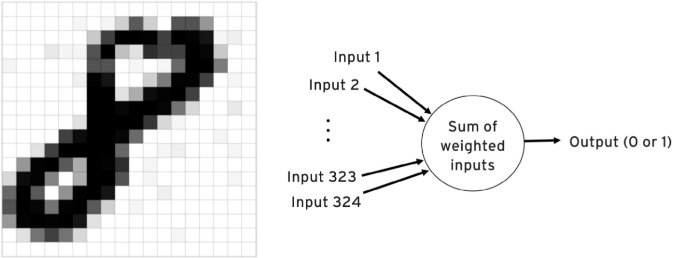

Illustration: M. Mitchell

Rosenblatt's perceptron

- Inputs (pixel intensities) with weights
- Nodes with activation levels from 0-1
- (Perhaps) 0-1 output based on a threshold

Word2vec

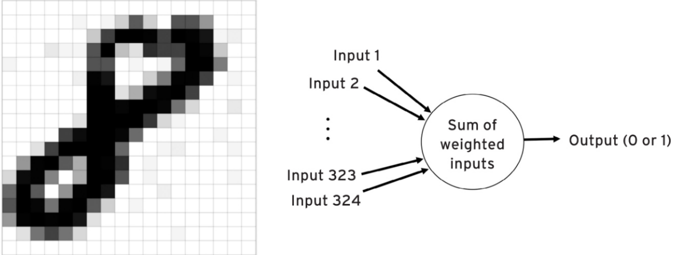

Illustration: M. Mitchell

Learning

- Start with random weights
- Test on a case:
    - If right, don't change weights.
    - If wrong, change weights a bit, with focus on the ones more responsible for the judgment:

$$w_j \leftarrow w_j = \overbrace{\eta}^{\text{learning rate}} \; (\underbrace{t}_{\text{correct output}} - \overbrace{y}^{\text{actual output}}) \; \underbrace{x_j}_{\text{actual input}}$$

Word2vec

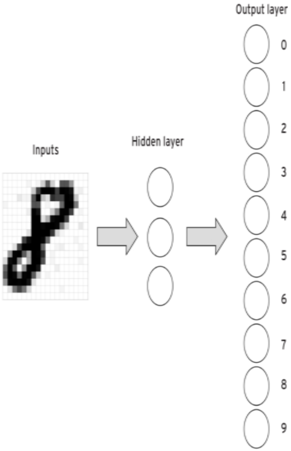

Illustration: M. Mitchell

- Each hidden unit takes a weighted sum of 324 inputs and passes on its activation level as input to outer layer units.
- Activation levels of outer layers are interpreted as network's levels of confidence in a classification problem.
- Learning: back-propagation (gradient descent: approximate the direction of steepest descent in the error surface w.r.t to weights, modify accordingly).

Word2vec

Distributional semantics

- "You shall know a word by the company it keeps" (John Firth, 1957)
- "the degree of semantic similarity between two linguistic expressions *A* and *B* is a function of the similarity of the linguistic contexts in which *A* and *B* can appear." (A. Lenci, 2008)

Word2vec

Distributional semantics

- "You shall know a word by the company it keeps" (John Firth, 1957)
- "the degree of semantic similarity between two linguistic expressions *A* and *B* is a function of the similarity of the linguistic contexts in which *A* and *B* can appear." (A. Lenci, 2008)

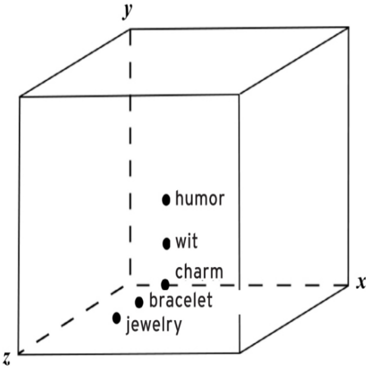

Illustration: M. Mitchell

Word2vec

Google and Mikolov

Efficient Estimation of Word Representation in Vector Space, 2013
Let's train a neural network and use vectors of weights!

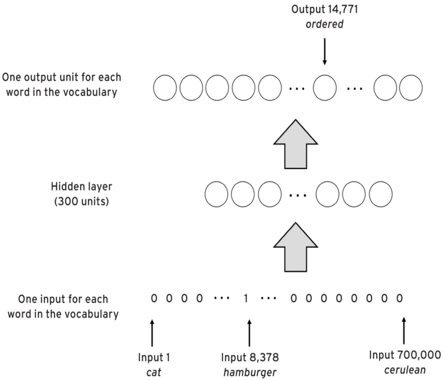

Illustration: M. Mitchell

Word2vec

Google and Mikolov

Efficient Estimation of Word Representation in Vector Space, 2013

Let's train a neural network and use vectors of weights!

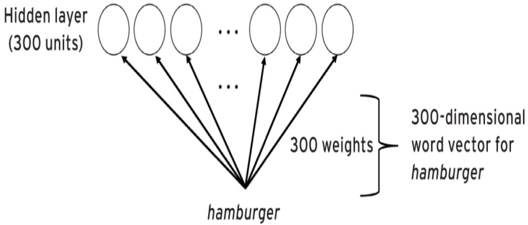

Illustration: M. Mitchell

Word2vec

Nearest words

- **philosophy**: philosophies, credo, ethos, principles, ethic, tenets, mantra, ideology, mindset, worldview
- **sandwich**: sandwiches, burger, chicken sandwich, cheeseburger, burrito, burgers, pizza, turkey sandwich, hamburger, burritos

Word2vec

Nearest words

- **philosophy**: philosophies, credo, ethos, principles, ethic, tenets, mantra, ideology, mindset, worldview
- **sandwich**: sandwiches, burger, chicken sandwich, cheeseburger, burrito, burgers, pizza, turkey sandwich, hamburger, burritos

Some similarities from *philosophy*

Logic (.47), Nietzsche (.32), Hegel (.32), analytic (.13), burger (.08), continental (.04), Russell (.04)