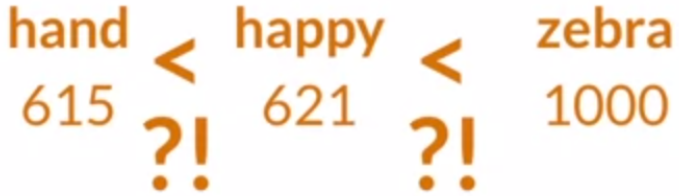
Natural Language Processing with Probabilistic Models: Word embeddings

Representing words:

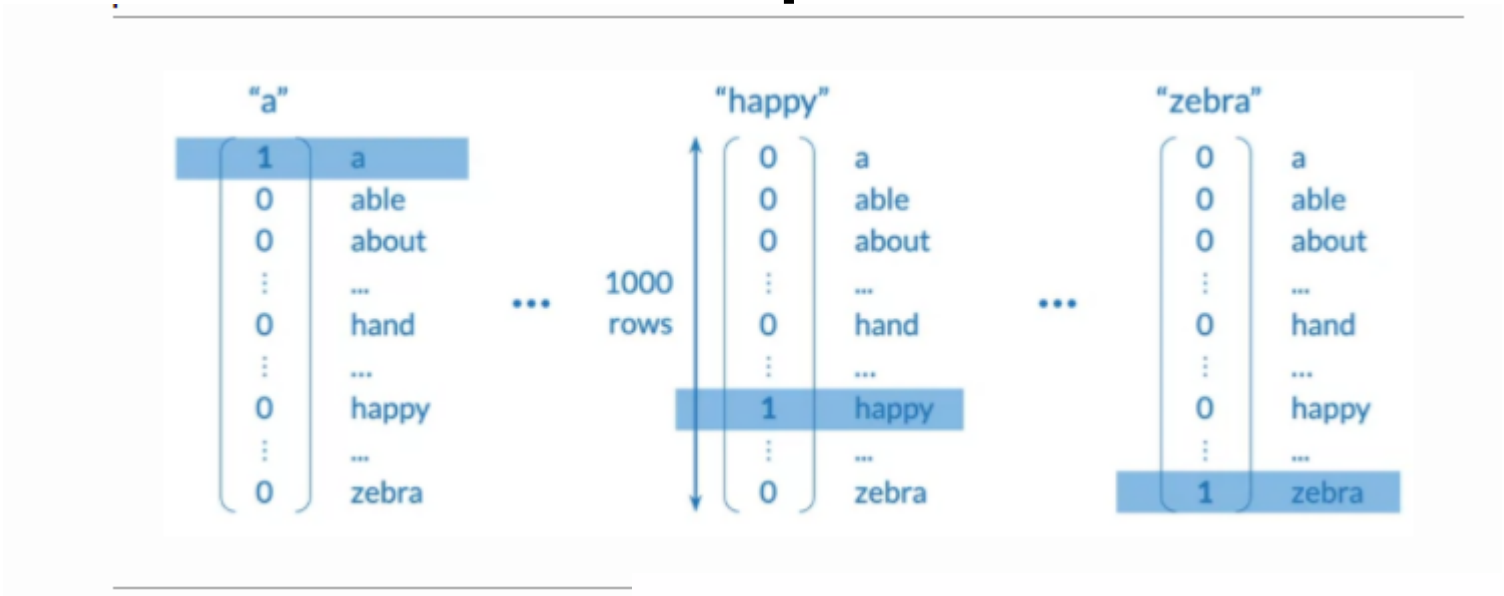
- The simplest way to represent words as numbers is for a given vocabulary to assign a unique integer to each word
  -

Word	Number
a	1
able	2
about	3
...	...
hand	615
...	...
happy	621
...	...
zebra	1000

- Although it's simple it has little semantic sense

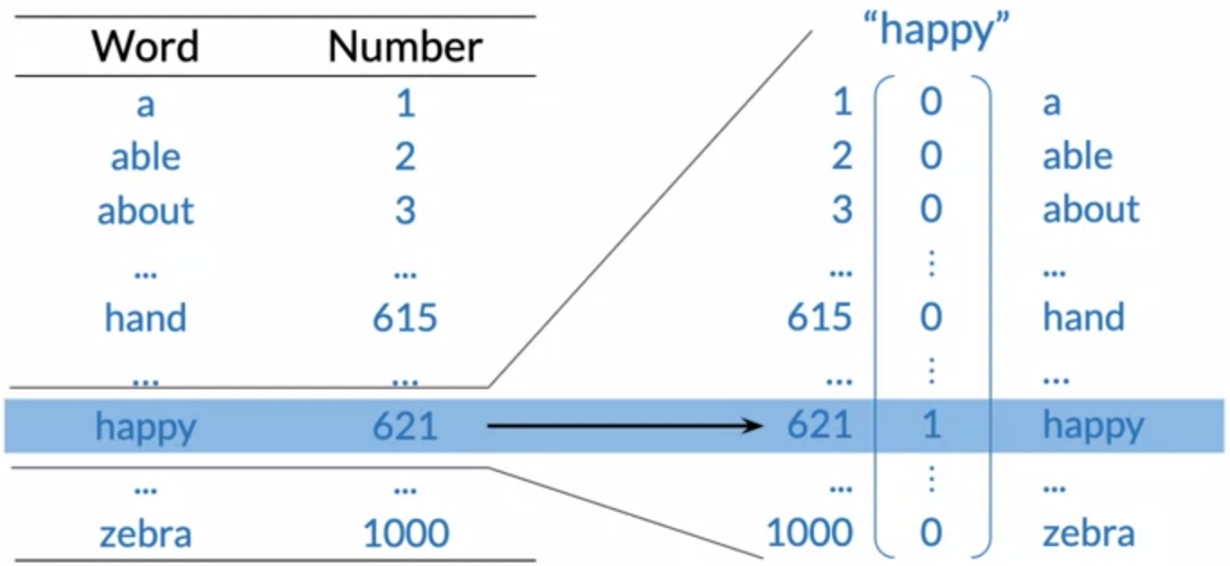


One hot vector representation



One-hot vectors

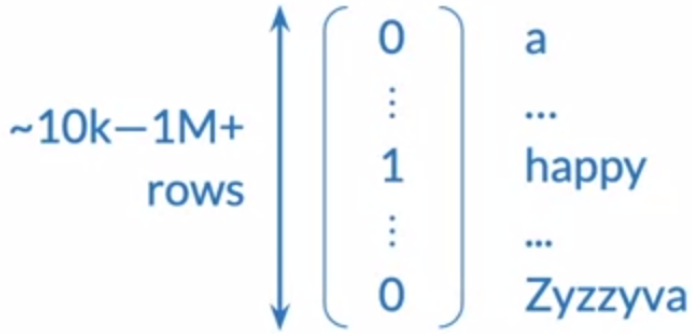
Word	Number
a	1
able	2
about	3
...	...
hand	615
...	...
happy	621
...	...
zebra	1000

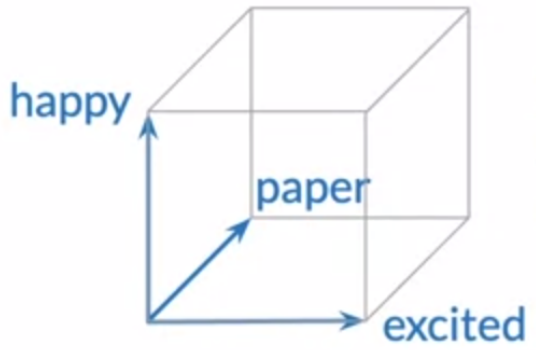


One hot vectors: Advantages and limitations,

One-hot vectors

+ Simple
+ No implied ordering
- Huge vectors
- No embedded meaning





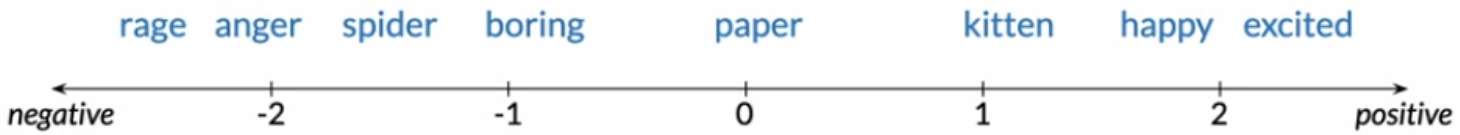
d(paper, excited)
= d(paper, happy)
= d(excited, happy)

it looks that paper is 'similar' to happy and not 'excited'.

And there the word embeddings come in

Word Embeddings

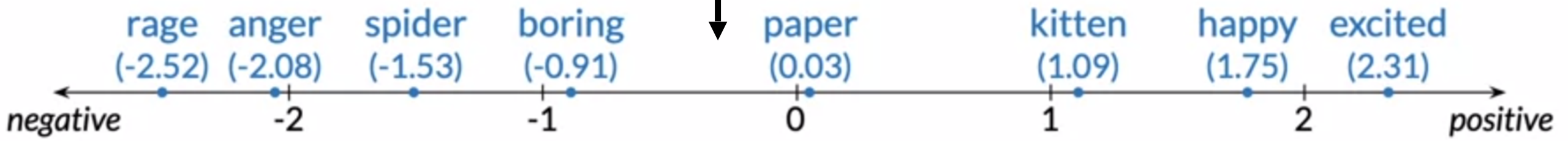
Meaning as vectors



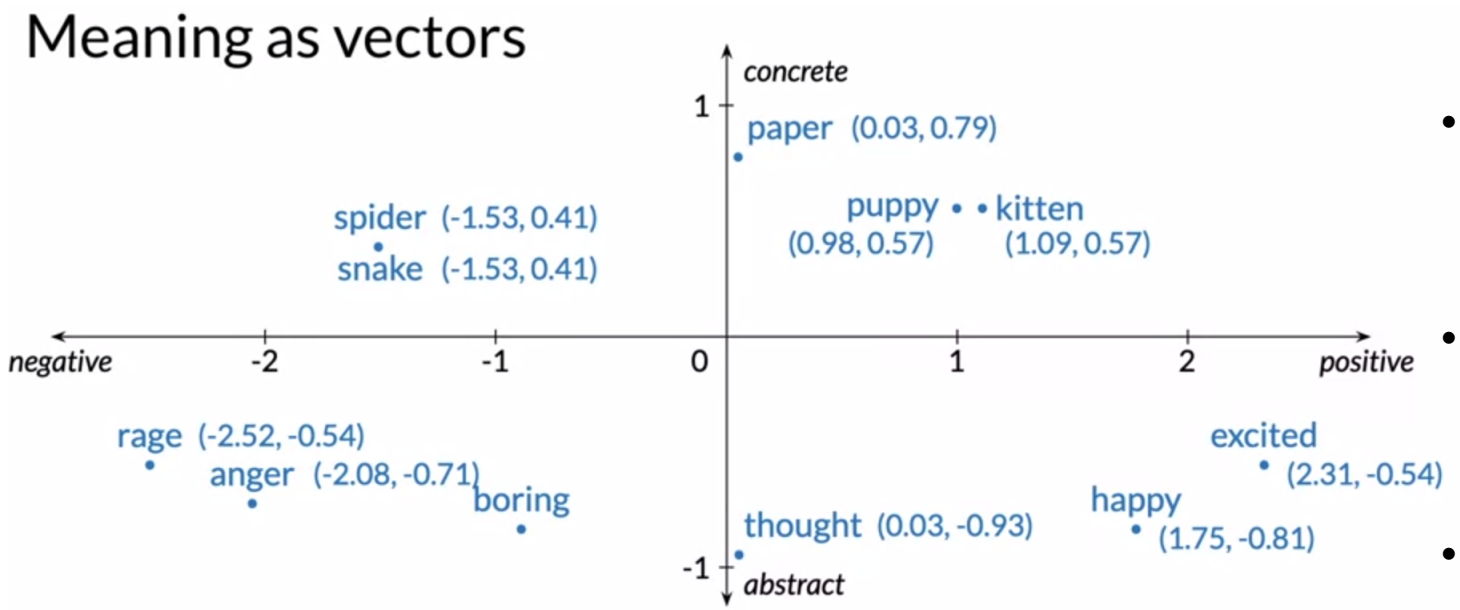
more negative words further to the left

more positive words further to the right

you could use decimal values than 0 and 1 which makes sense if words are similar like: 'happy' and 'excited'



Word Embeddings



- What you created just now is an example of a word embedding.
- Word embeddings represent words in a vector form that's both has a relatively low dimension saying the hundreds to load thousands.
- Making it practical for calculations and carries the meaning of words making it possible to determine how semantically closed words are. I
- You will visualize such similarities as part of this week's assignment.
- It also makes it possible to work out analogies, such as finding the missing word in Paris is to France as Rome is to?

Word Embeddings

Terminology

integers

word vectors

one-hot vectors

word embedding vectors

"word vectors"

word embeddings

Word Embeddings

Summary

- Words as integers
- Words as vectors
  - One-hot vectors
  - Word embedding vectors
- Benefits of word embeddings for NLP

Word Embedding Methods (Advanced methods using deep neural networks)

Advanced word embedding methods

Deep learning, contextual embeddings

different embeddings for the same word depending upon the context

- BERT (Google, 2018)
- ELMo (Allen Institute for AI, 2018)
- GPT-2 (OpenAI, 2018)

} Tunable pre-trained models available

download a pre-trained corpus and fine tune it