Word embeddings

NLP in one day

KING'S College LONDON

NIHR | Maudsley Biomedical Research Centre

Representing words as vectors

- We need to pass numbers to our NLP models
- One-hot is a simple word-space vector representation. Words are represented by a vector encoding their position in an ordered vocabulary

```
aardvark        [1, 0, 0, 0, 0, …, 0, 0]
abacus          [0, 1, 0, 0, 0, …, 0, 0]
...             [...]
zumba           [0, 0, 0, 0, 0, …, 1, 0]
zygote          [0, 0, 0, 0, 0, …, 0, 1]
```

- As well as being necessary to represent our words numerically, it is also a step along the path of finding some abstraction of word meaning
- Alternatively, we could encode as the integer position in the vocabulary

```
aardvark        0
abacus          1
...             ...
zumba           n-1
zygote          n
```

Representing words as vectors

- Such a vector representation does not really encode meaning
- It is also high dimensional and sparse
- Can we encode meaning such a vector representation?
- Can we derive a low dimensional model of words?

- What about the context vectors we have looked at? Would they be an improvement?

Intuition

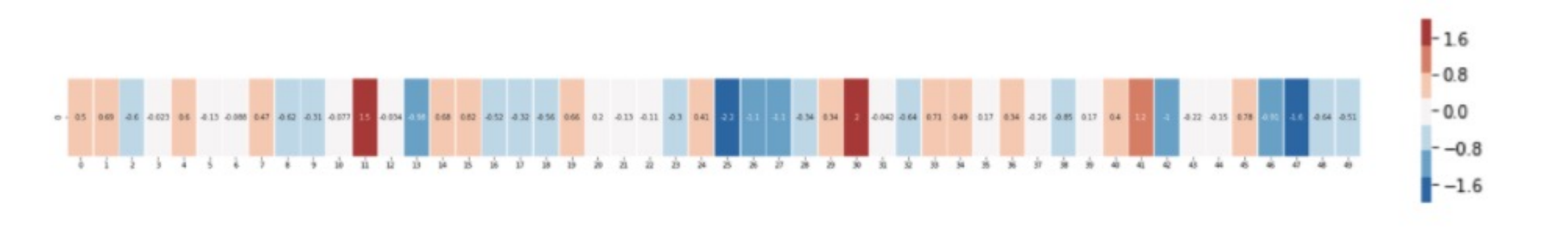
Imagine a vector for the word “king”, (GloVe based vector, trained on Wikipedia):

```
[ 0.50451 , 0.68607 , -0.59517 , -0.022801, 0.60046 , -0.13498 , -0.08813 , 0.47377 , -0.61798 , -0.31012 ,
-0.076666, 1.493 , -0.034189, -0.98173 , 0.68229 , 0.81722 , -0.51874 , -0.31503 , -0.55809 , 0.66421 , 0.1961
, -0.13495 , -0.11476 , -0.30344 , 0.41177 , -2.223 , -1.0756 , -1.0783 , -0.34354 , 0.33505 , 1.9927 ,
-0.04234 , -0.64319 , 0.71125 , 0.49159 , 0.16754 , 0.34344 , -0.25663 , -0.8523 , 0.1661 , 0.40102 , 1.1685 ,
-1.0137 , -0.21585 , -0.15155 , 0.78321 , -0.91241 , -1.6106 , -0.64426 , -0.51042 ]
```

Example from Jay Alammar, The illustrated Word2Vec:
https://jalammar.github.io/illustrated-word2vec/

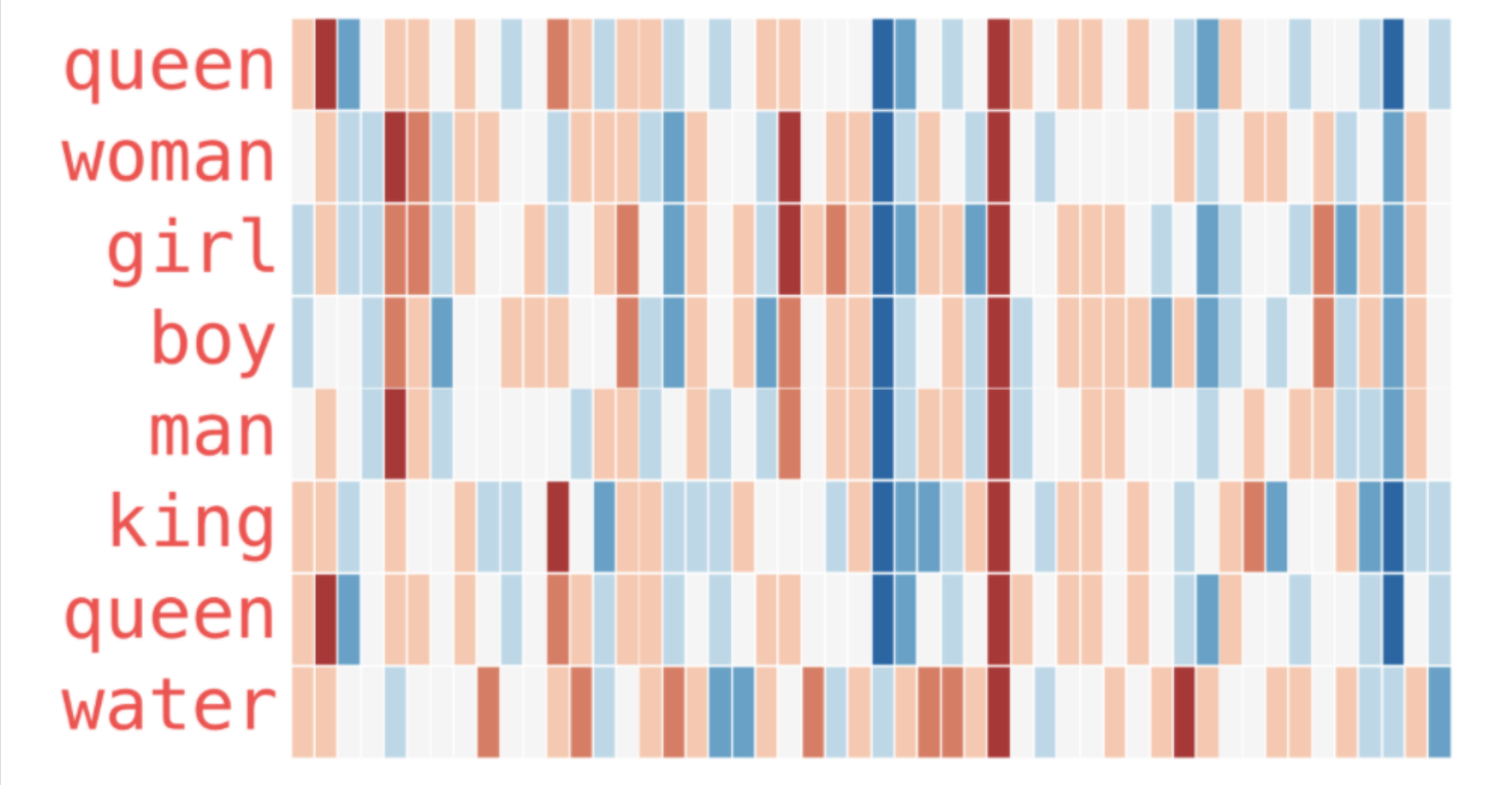
Intuition

We can visualise this as bands of different colours and intensities:

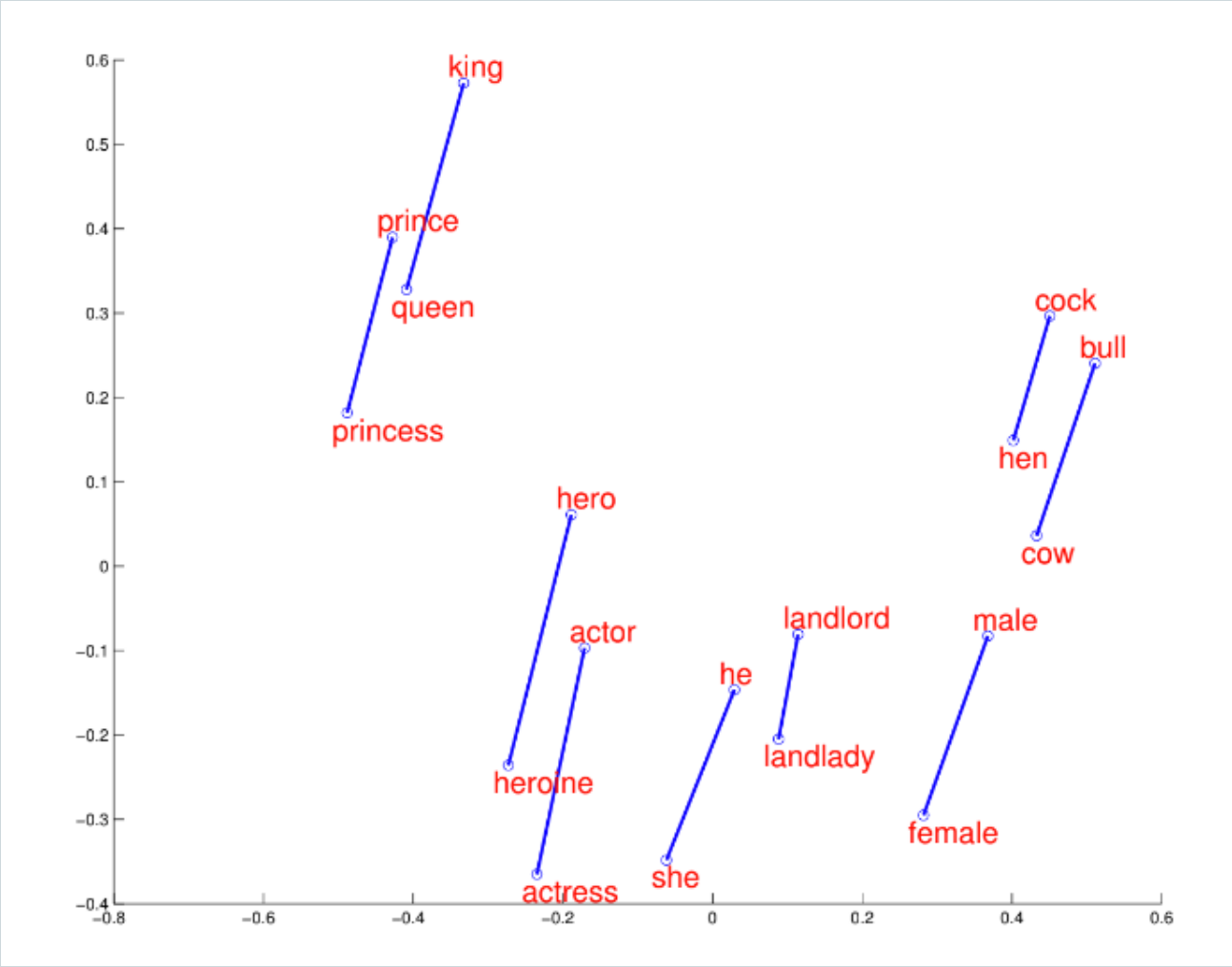


Intuition

Compare this word vector to vectors for other words:

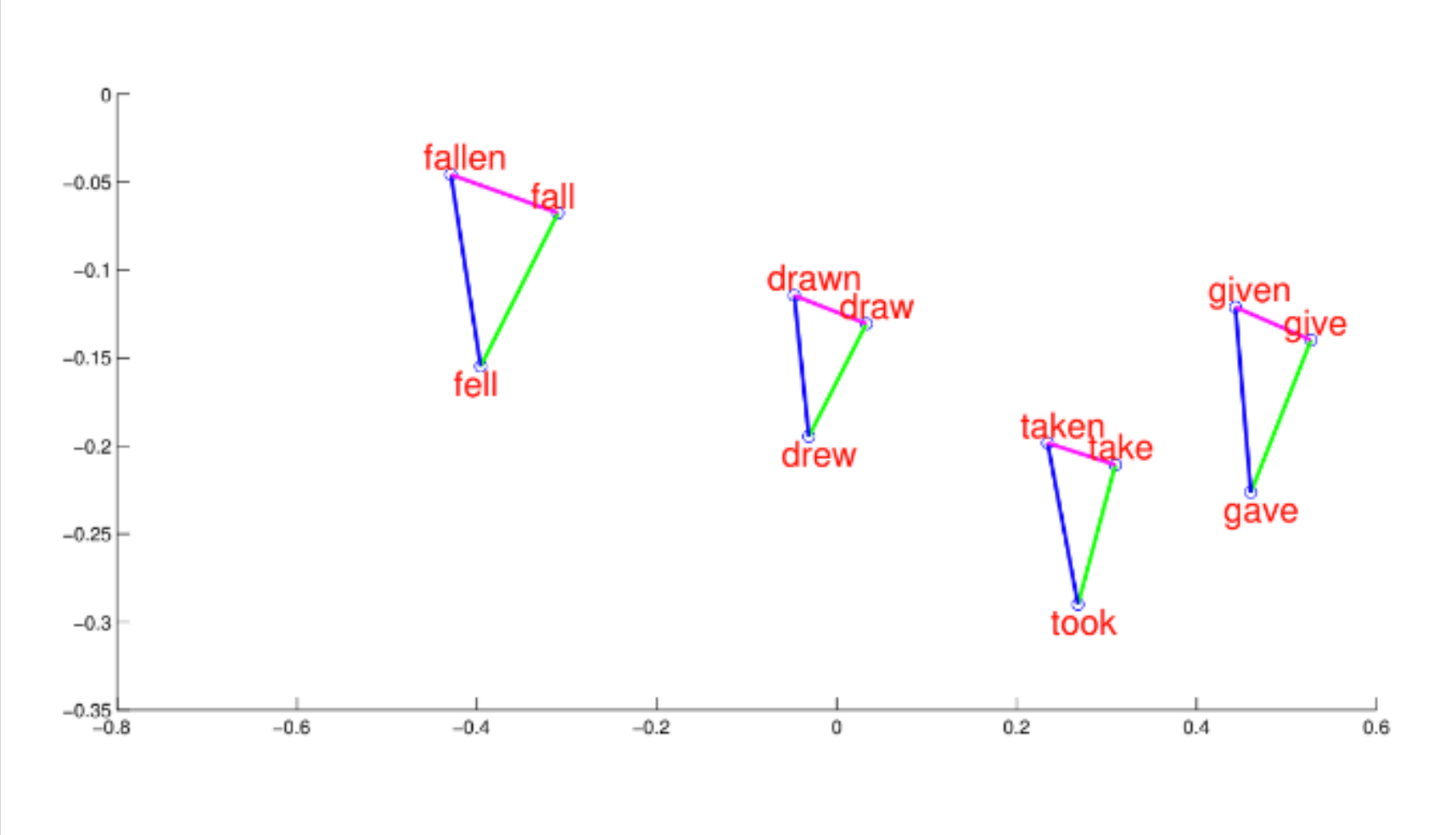


Visualisation



2D projection from Mikolov et al, Google Research, NIPS 2013

Visualisation



2D projection from Mikolov et al, Google Research, NIPS 2013

Thank you

angus.roberts@kcl.ac.uk
https://www.kcl.ac.uk/people/angus-roberts