Properties and Application of Word Embeddings

Sandeep Soni

10/18/2023

From last class

- Distributional hypothesis: Words used in similar contexts have similar meanings (and therefore similar representations)

- We can learn word vectors from a large corpus of text using word2vec or GloVe.

x	y
wine	cold
wine	sweet
wine	spirit
wine	drink

x	y
wine	cold
wine	sweet
wine	spirit
wine	drink

x	y
wine	cold
wine	sweet
wine	spirit
wine	drink

wine

x

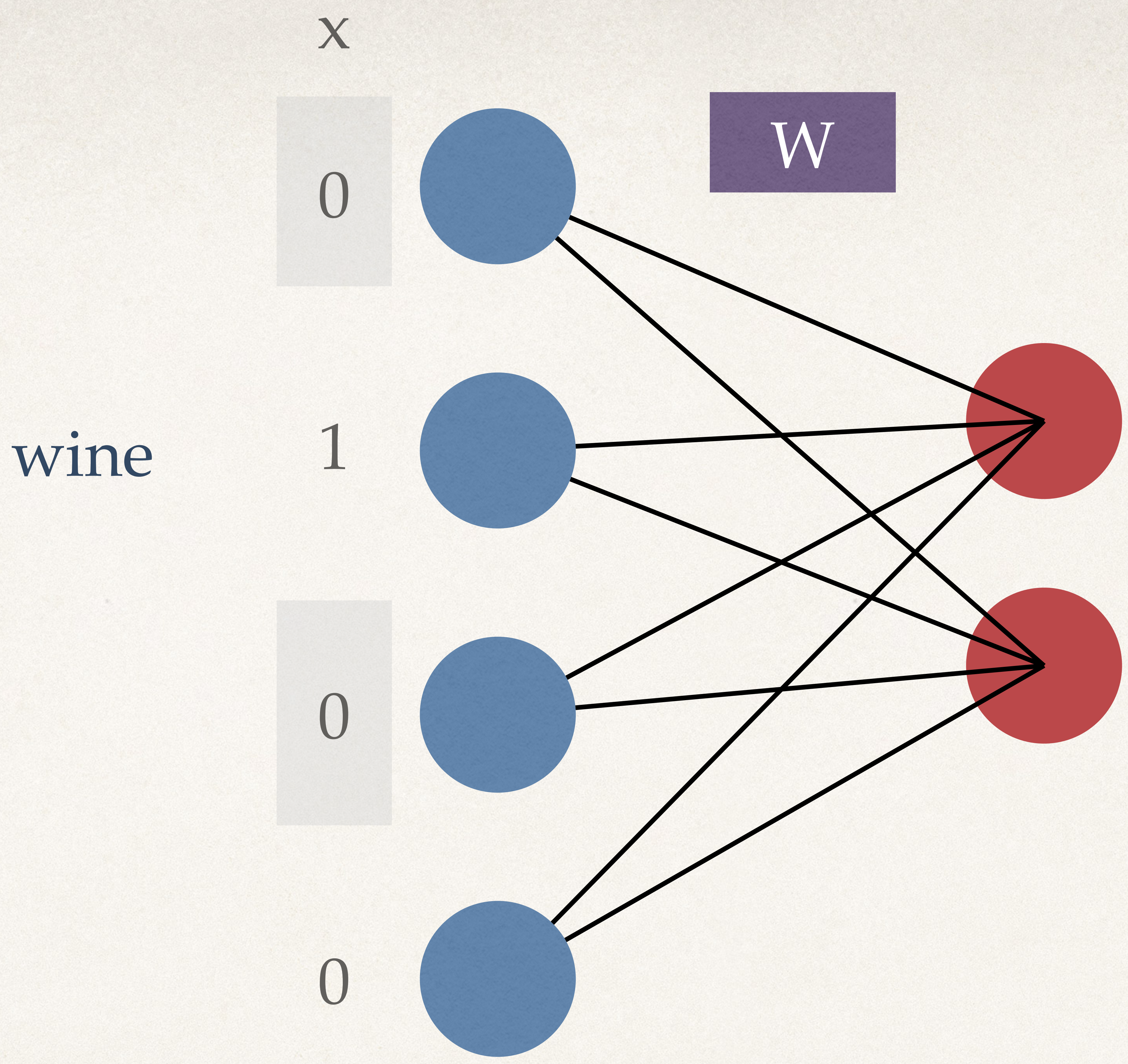
0

wine 1

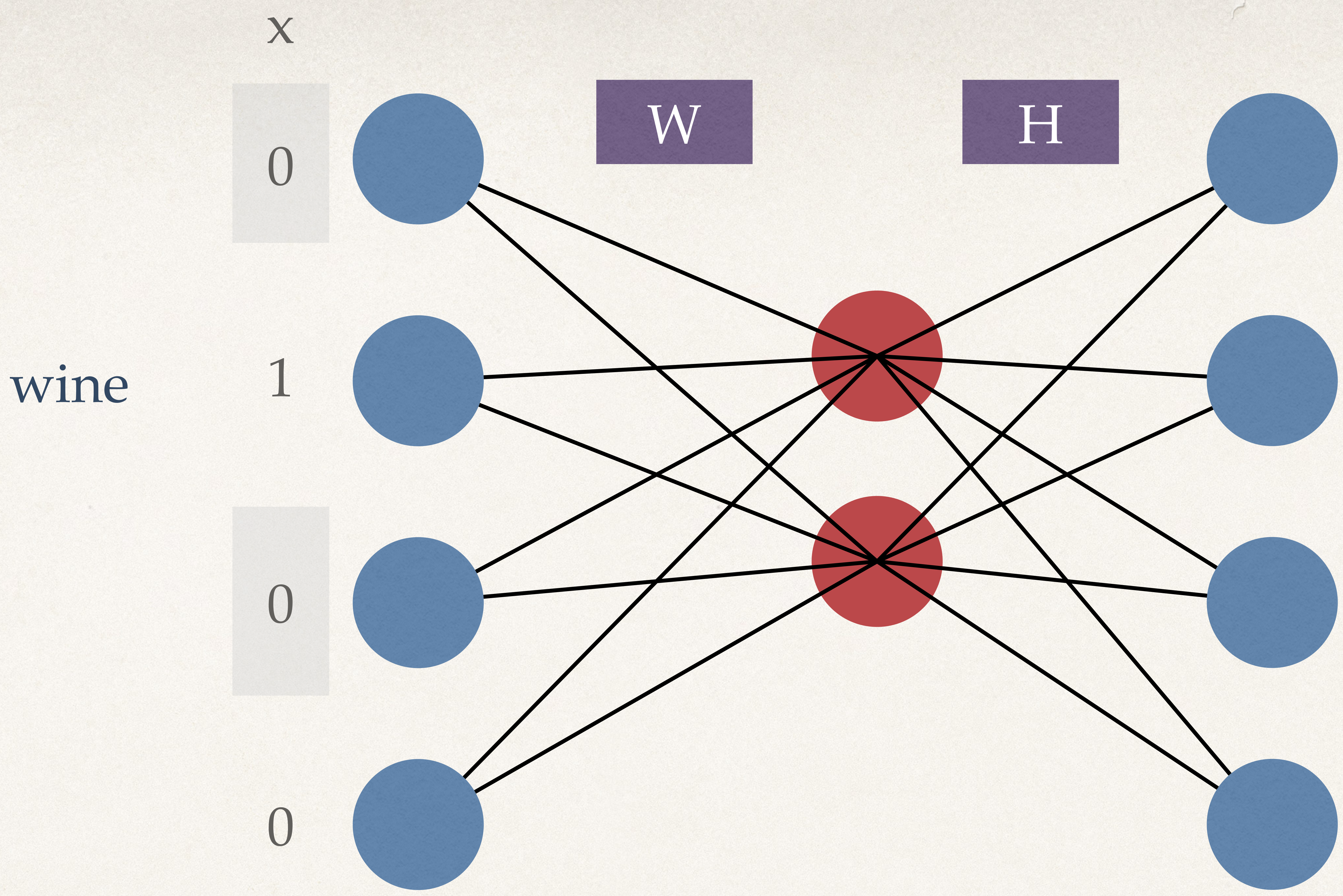
0

0

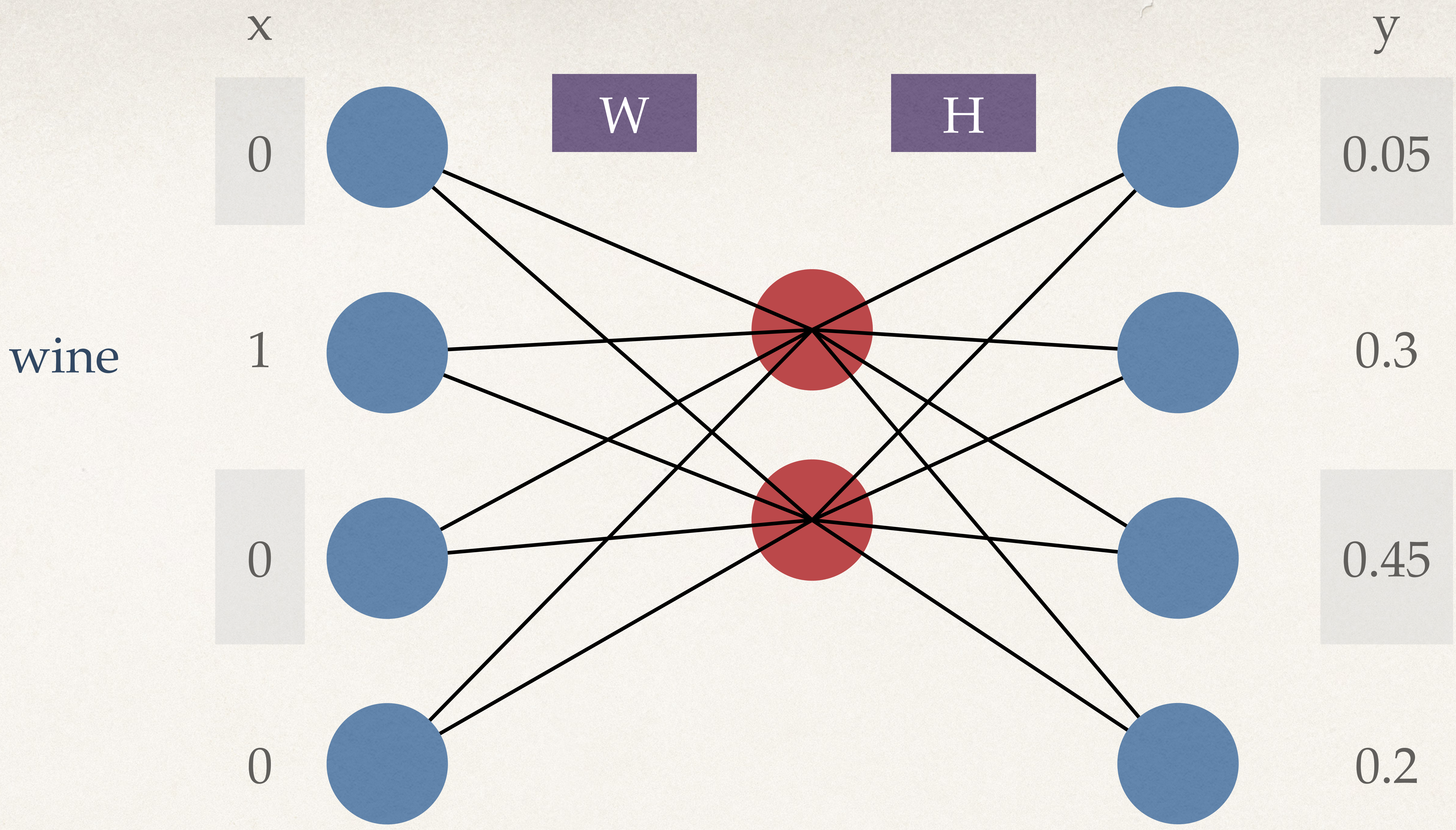
x	y
wine	cold
wine	sweet
wine	spirit
wine	drink



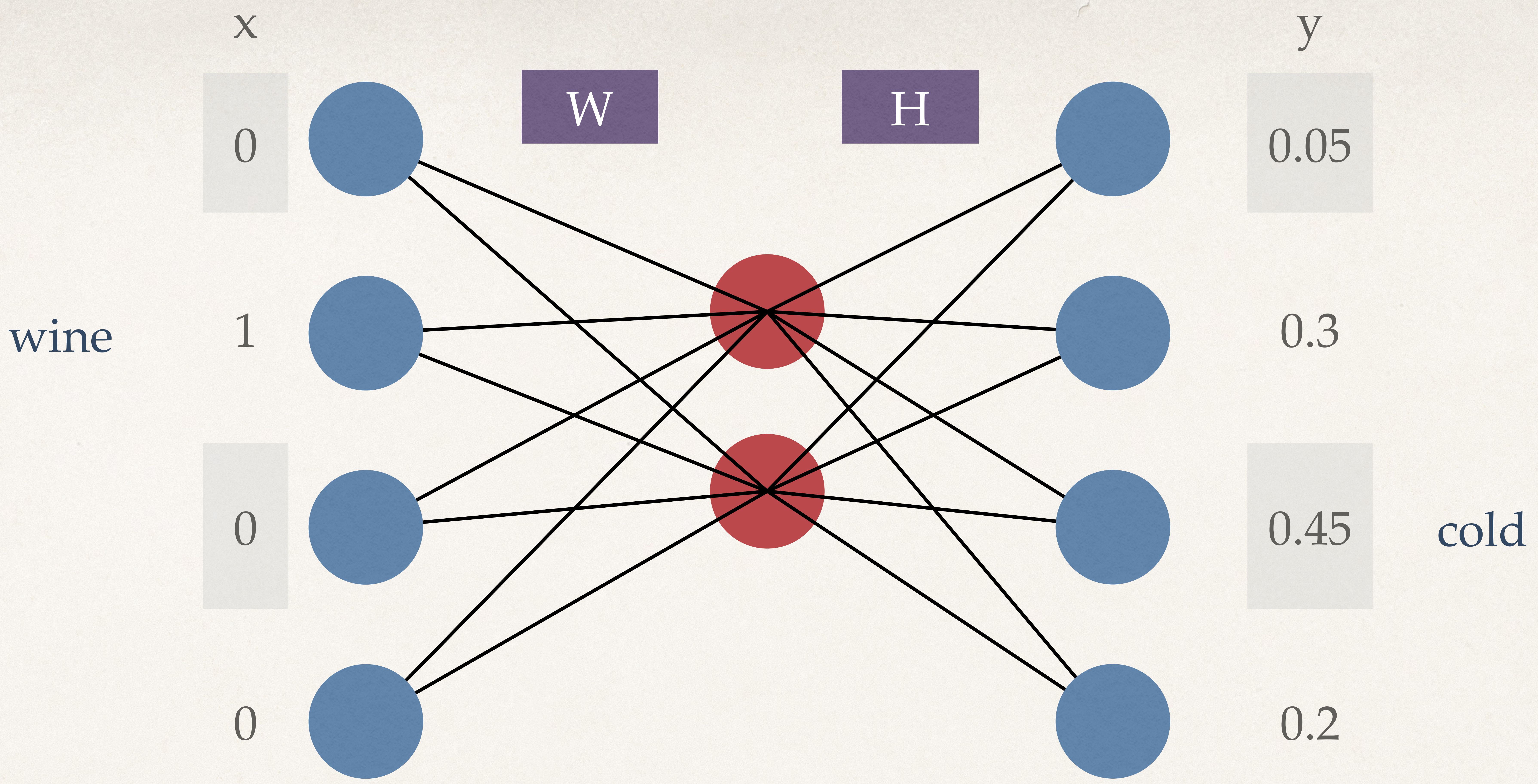
x	y
wine	cold
wine	sweet
wine	spirit
wine	drink



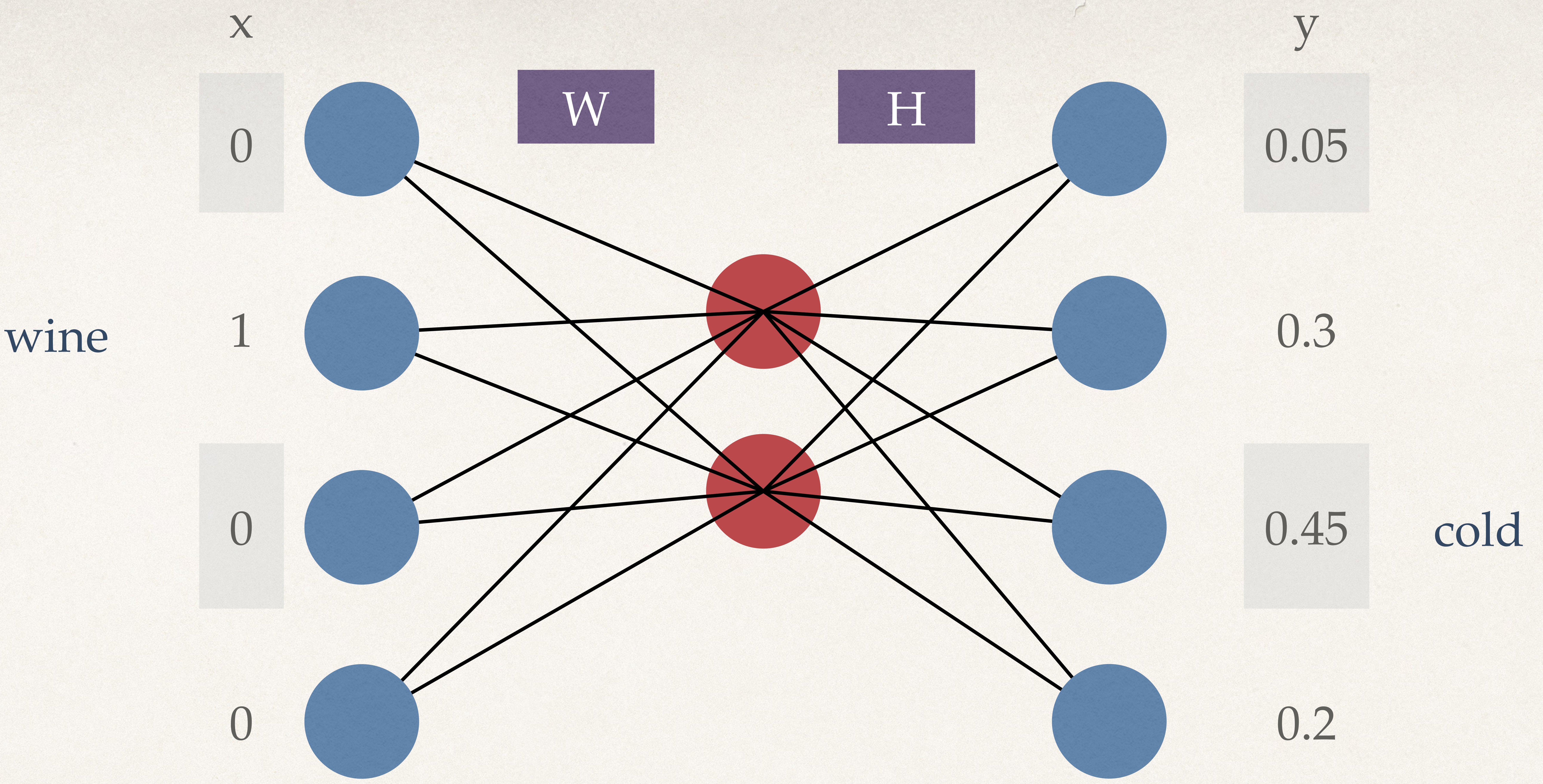
x	y
wine	cold
wine	sweet
wine	spirit
wine	drink



x	y
wine	cold
wine	sweet
wine	spirit
wine	drink



x	y
wine	cold
wine	sweet
wine	spirit
wine	drink

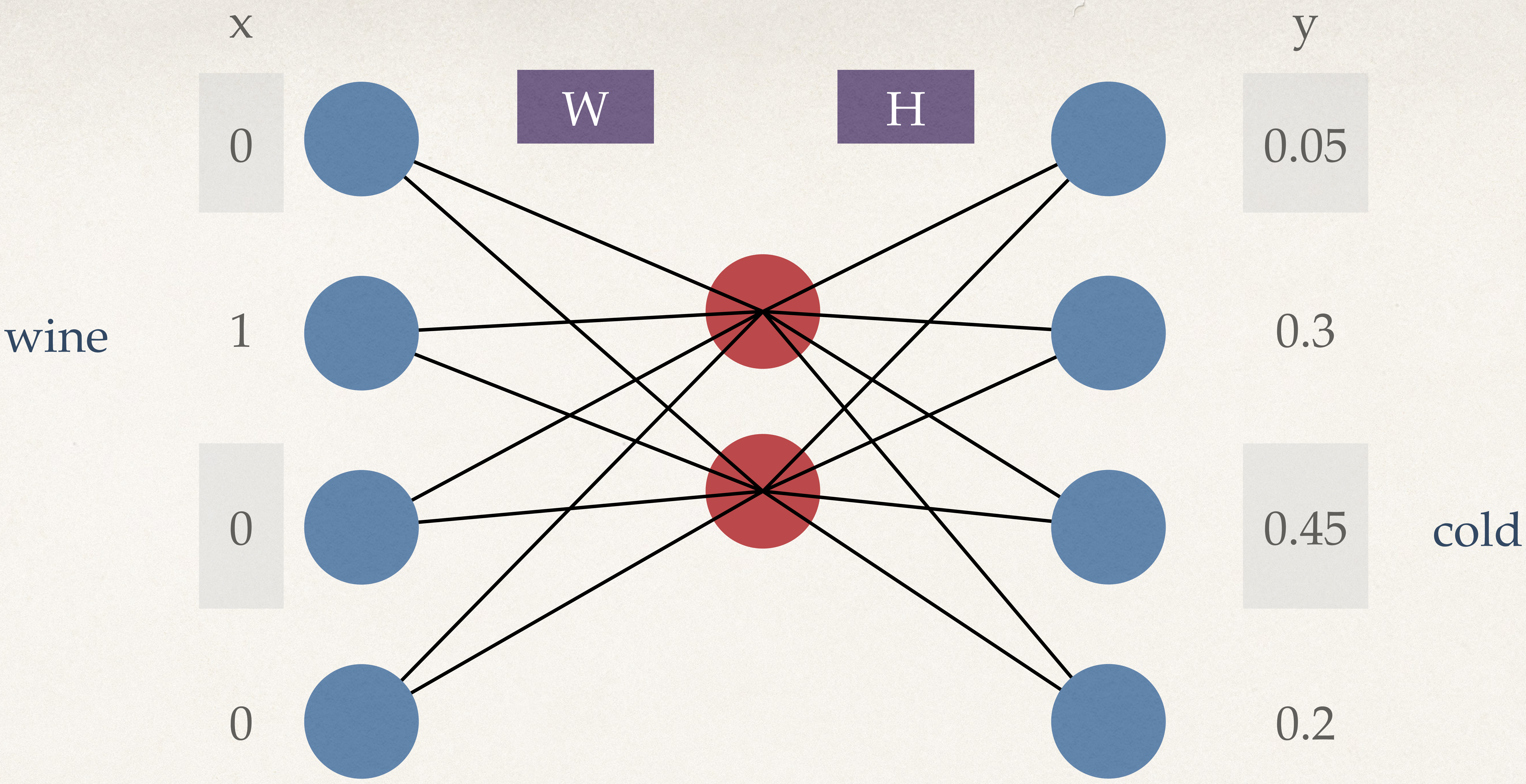


x	y
wine	cold
wine	sweet
wine	spirit
wine	drink

Word embeddings as columns

W

-0.3	1.2	0.5	-0.6
0.2	0.9	0.1	-0.4



x	y
wine	cold
wine	sweet
wine	spirit
wine	drink

Word embeddings as columns

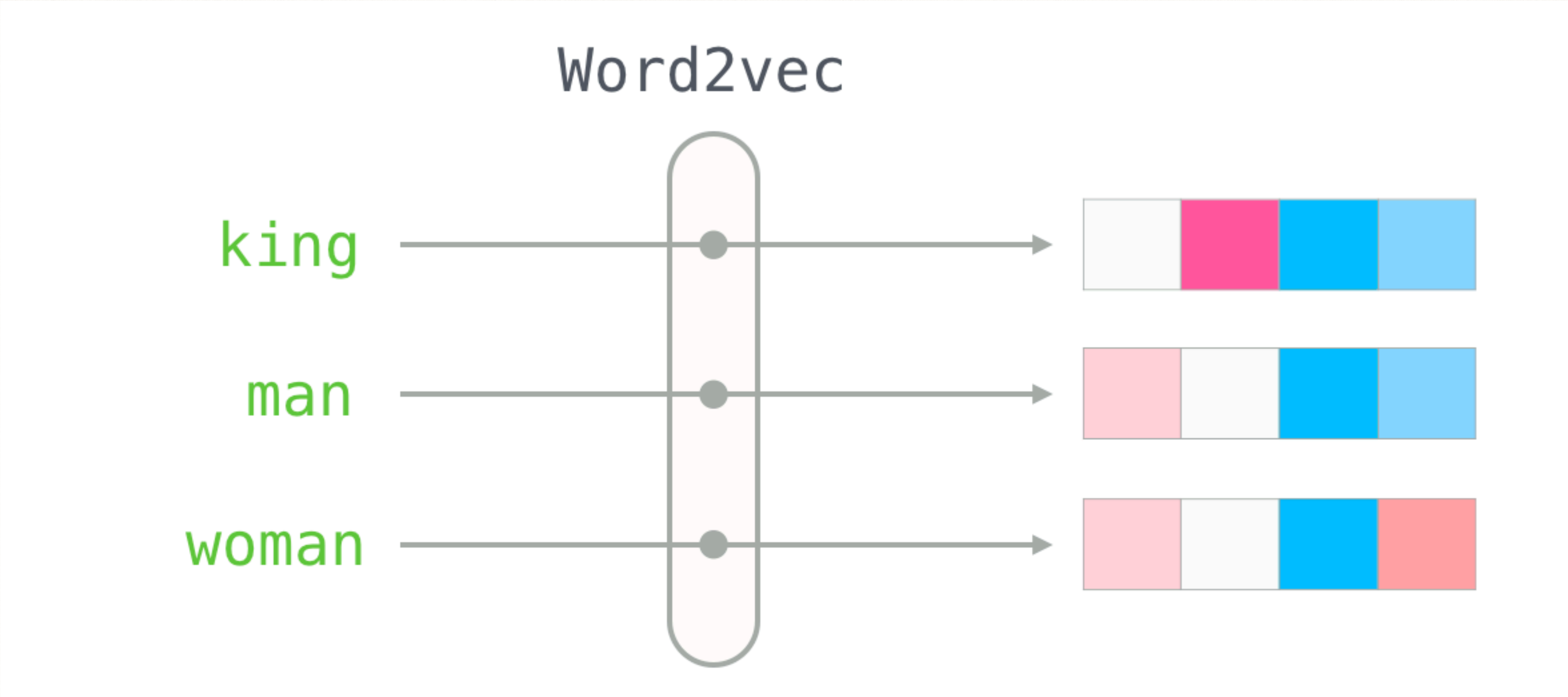
W

-0.3	1.2	0.5	-0.6
0.2	0.9	0.1	-0.4

Context embeddings as rows

H

0.1	-0.4
0.4	-0.5
0.3	-0.1
0.2	0.1



http://jalammar.github.io/illustrated-word2vec/

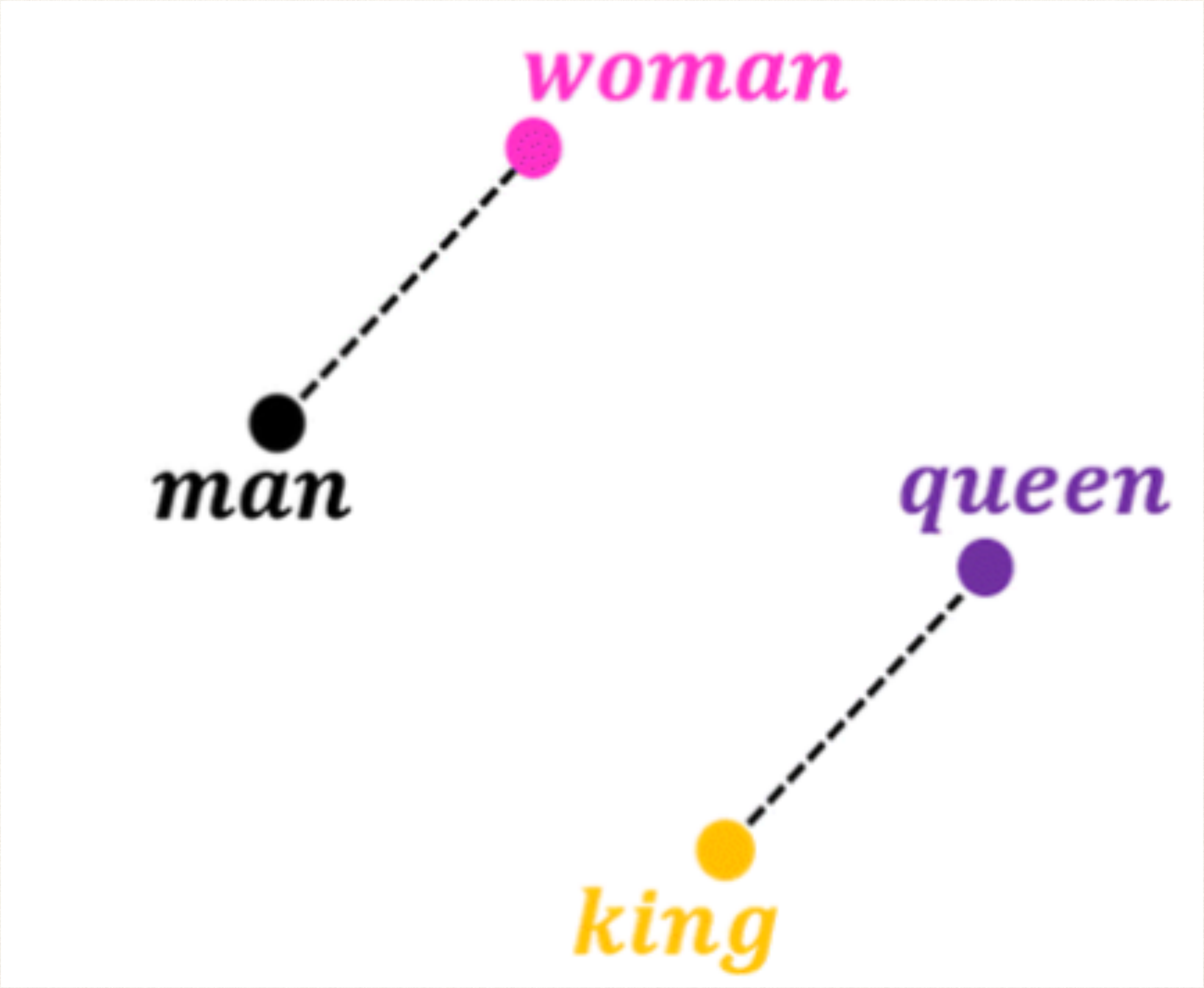
Question for the day

“What do word embeddings encode and what can we do with them?”

Geometry

man:woman::king:queen

v("man")-v("woman")+v("king") ≈ v("queen")



Mikolov et. al. 2013 Linguistic Regularities in Continuous Space Word Representations

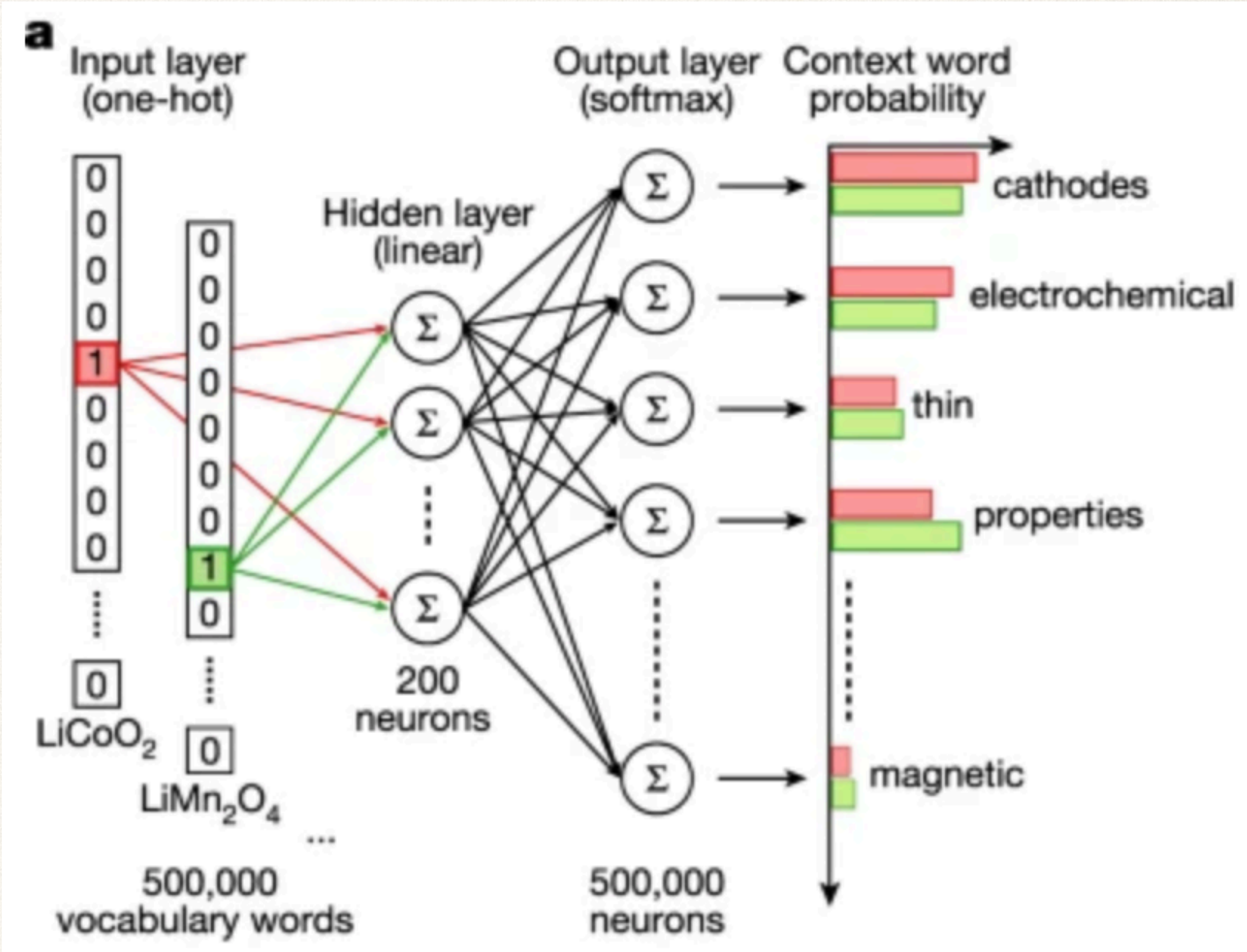
Prediction

$$\text{ferromagnetic} - \text{NiFe} + \text{IrMn} \approx \text{antiferromagnetic}$$

Oxides $Zr - ZrO_2 \approx Cr - Cr_2O_3 \approx Ni - NiO$

Structure $Zr - HCP \approx Cr - BCC \approx Ni - FCC$

Embeddings can be used to construct knowledge bases that can lead to new discoveries



Tshitoyan, Vahe, et al. "Unsupervised word embeddings capture latent knowledge from materials science literature." *Nature* 571.7763 (2019): 95-98.

Prediction

Are Word Embedding-based Features Useful for Sarcasm Detection?

Aditya Joshi[1,2,3] **Vaibhav Tripathi**[1] **Kevin Patel**[1]
Pushpak Bhattacharyya[1] **Mark Carman**[2]
[1]Indian Institute of Technology Bombay, India
[2]Monash University, Australia
[3]IITB-Monash Research Academy, India
{adityaj,kevin.patel,pb}@cse.iitb.ac.in, mark.carman@monash.edu

	Word2Vec	LSA	GloVe	Dep. Wt.
+S	0.835	0.86	0.918	**0.978**
+WS	**1.411**	0.255	0.192	1.372
+S+WS	**1.182**	0.24	0.845	0.795

Table 4: Average gain in F-Scores obtained by using intersection of th four word embeddings, for three word embedding feature-types, aug mented to four prior works; Dep. Wt. indicates vectors learned fron dependency-based weights

Word Embedding	Average F-score Gain
LSA	0.452
Glove	0.651
Dependency	1.048
Word2Vec	1.143

Table 5: Average gain in F-scores for the four types of word embed dings; These values are computed for a subset of these embedding consisting of words common to all four

man:woman::king:?

man:woman::waiter:?

man:woman::doctor:?

`man:woman::king:?`	queen	okay
`man:woman::waiter:?`	waitress	okay
`man:woman::doctor:?`	nurse	huh

REPORT

COGNITIVE SCIENCE

Semantics derived automatically from language corpora contain human-like biases

Aylin Caliskan,[1*] Joanna J. Bryson,[1,2*] Arvind Narayanan[1*]

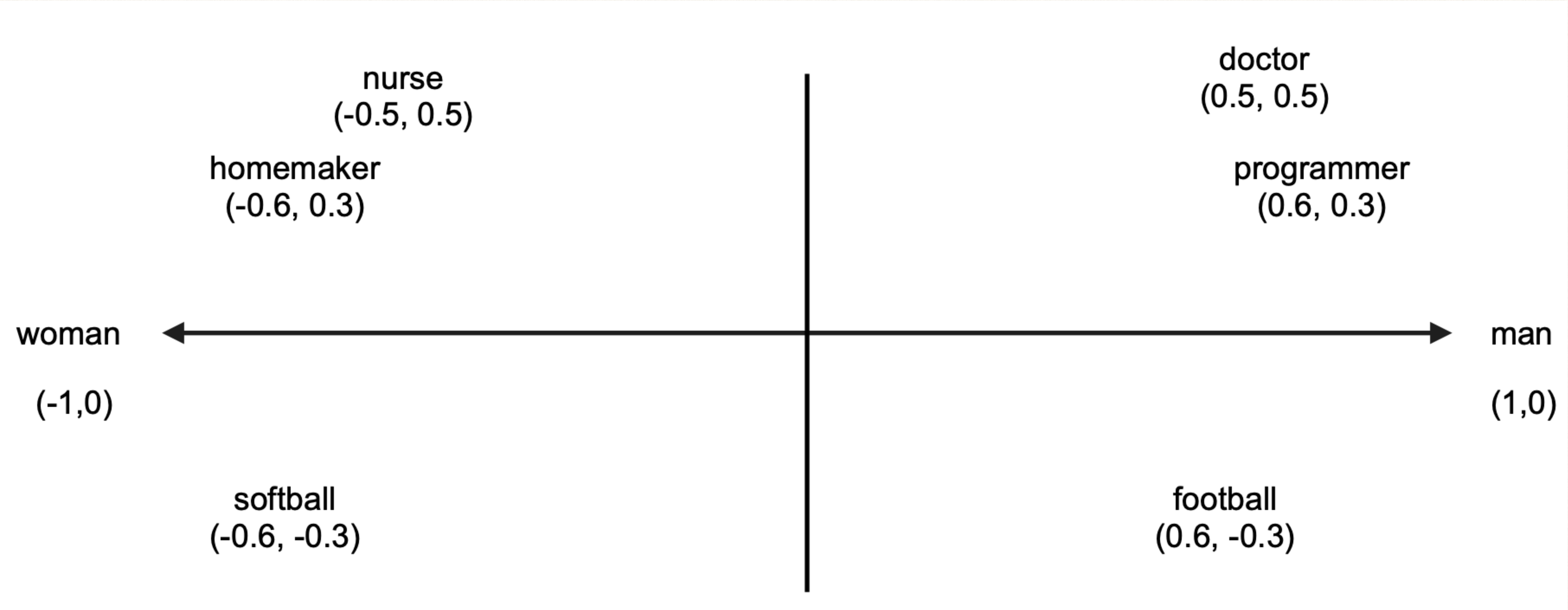
Machine learning is a means to derive artificial intelligence by discovering patterns in existing data. Here, we show that applying machine learning to ordinary human language results in human-like semantic biases. We replicated a spectrum of known biases, as measured by the Implicit Association Test, using a widely used, purely statistical machine-learning model trained on a standard corpus of text from the World Wide Web. Our results indicate that text corpora contain recoverable and accurate imprints of our historic biases, whether morally neutral as toward insects or flowers, problematic as toward race or gender, or even simply veridical, reflecting the status quo distribution of gender with respect to careers or first names. Our methods hold promise for identifying and addressing sources of bias in culture, including technology.

Bias

- Allocation harms: Systems should not allocate resources to groups unfairly
- Representational harms: Systems should not misrepresent groups unfairly

- Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of “Bias” in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5476, Online. Association for Computational Linguistics.

Bias



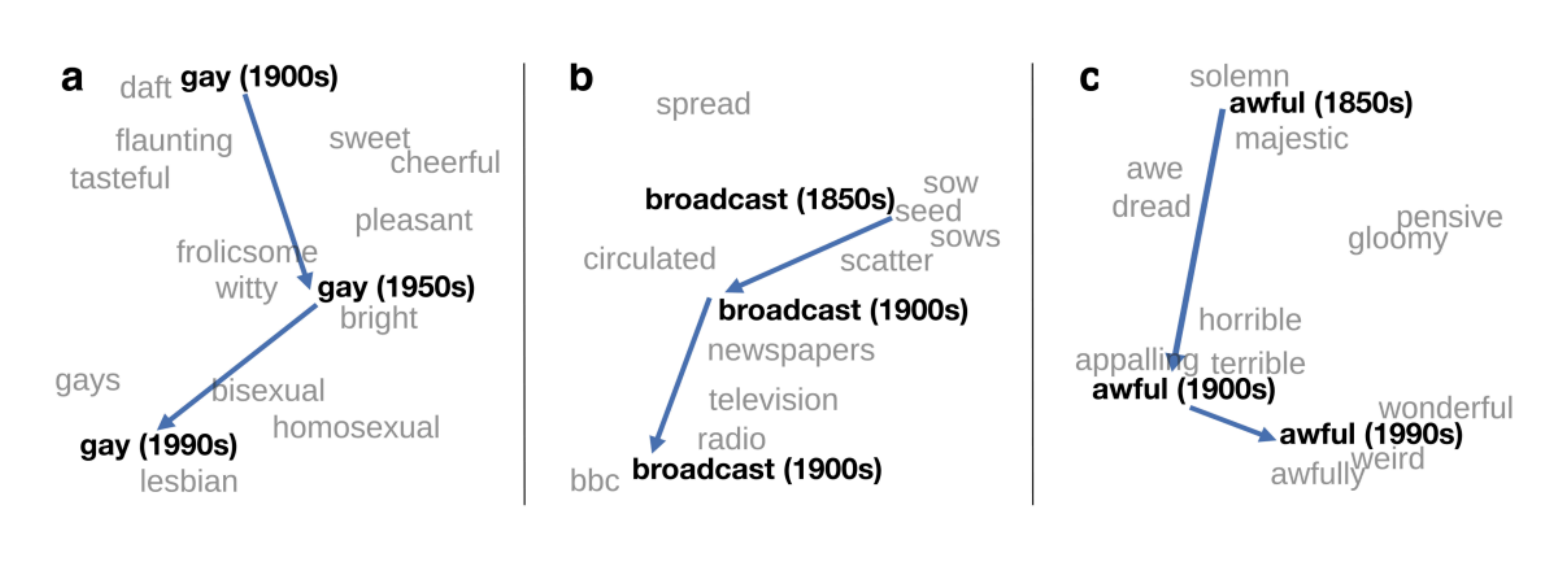
nurse
(-0.5, 0.5)
doctor
(0.5, 0.5)
homemaker
(-0.6, 0.3)
programmer
(0.6, 0.3)
woman
(-1,0)
man
(1,0)
softball
(-0.6, -0.3)
football
(0.6, -0.3)

My boy's *hella* smart

My boy's *wicked* smart

My boy's *very* smart

- David Bamman, Chris Dyer, and Noah A. Smith. 2014. Distributed Representations of Geographically Situated Language. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 828–834, Baltimore, Maryland. Association for Computational Linguistics.



- William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1489–1501, Berlin, Germany. Association for Computational Linguistics.

Change and Variation

- Since language is situational, one can learn embeddings that depend on time, geography or other social contexts

In class

- Word2Vec demo