Word2Vec

Word Embeddings

- **Goal**: find a vector representation of each word which captures a “meaning” of this word

- These vectors are called “**word embeddings**”

Distributional Hypothesis

Words that are used and occur in the same contexts tend to have similar meanings.

The underlying idea is that
"a word is characterized by the company it keeps"

Distributional Hypothesis

- My name is John
- My name is Mary
- My name is Jack
- ….

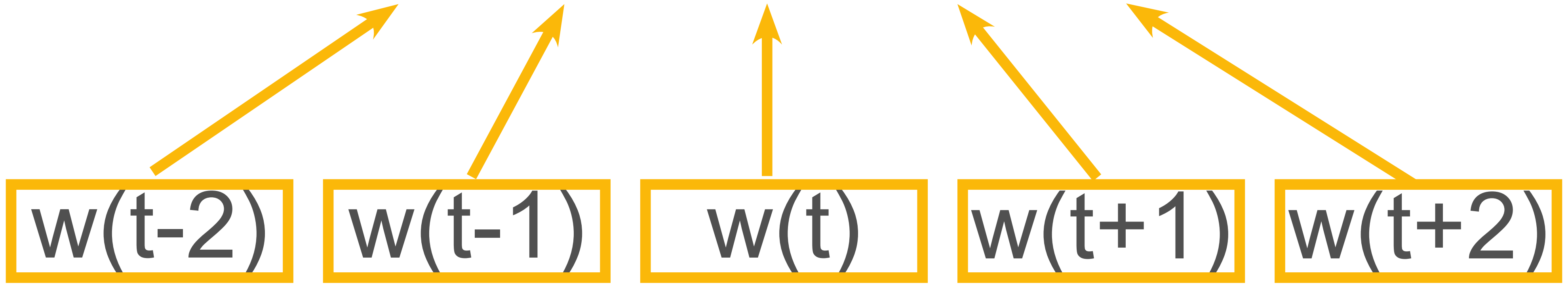
John, Mary, Jack – are **given names**

Inventors of Word2Vec

- Mikolov, T., Corrado, G., Chen, K., & Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of ICLR 2013.

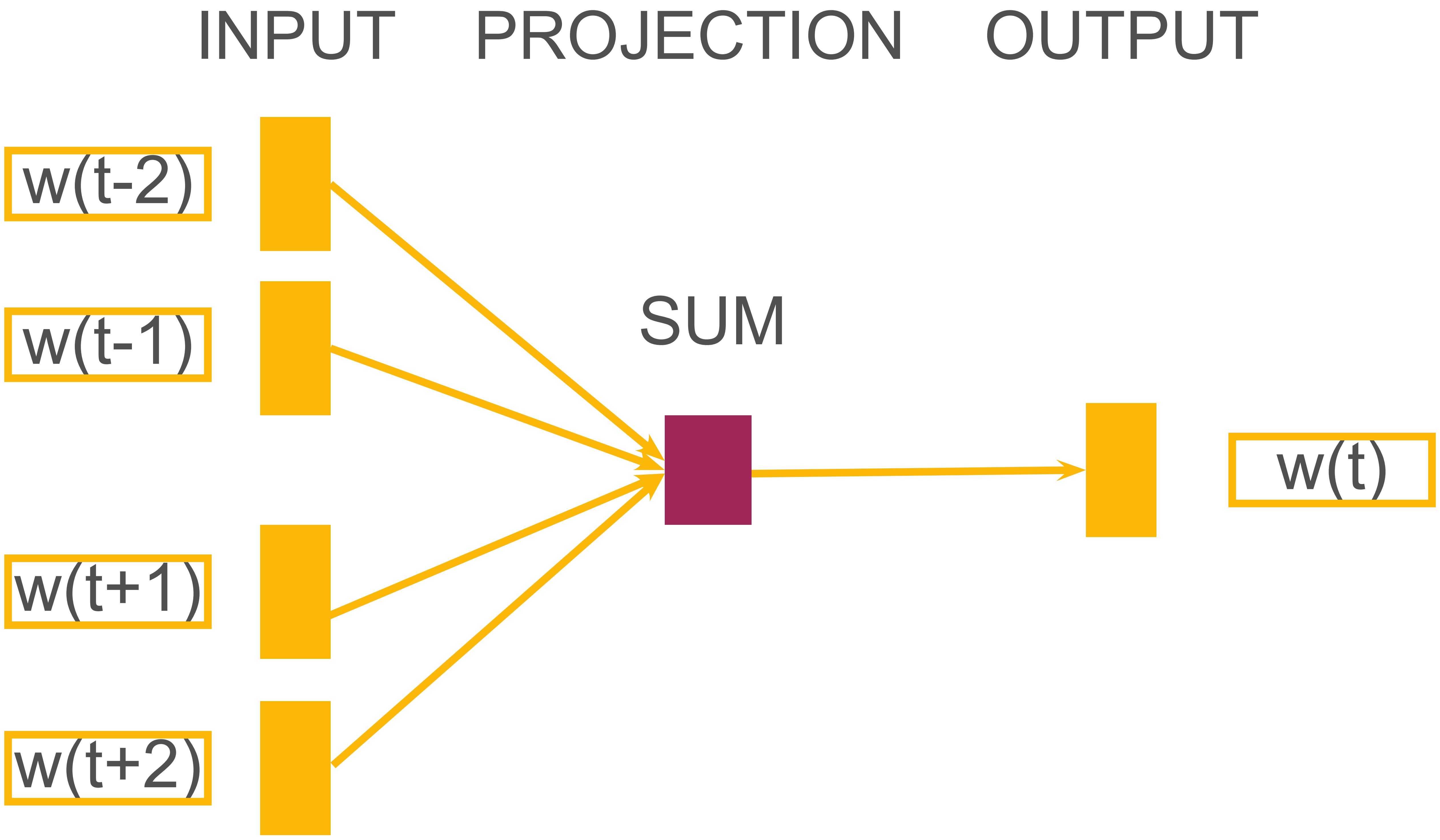
- **Mikolov, T.** , Sutskever I., Chen K., Corrado G., Dean J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In NIPS Deep Learning Workshop.

“The forest of oak trees on the mountain …”

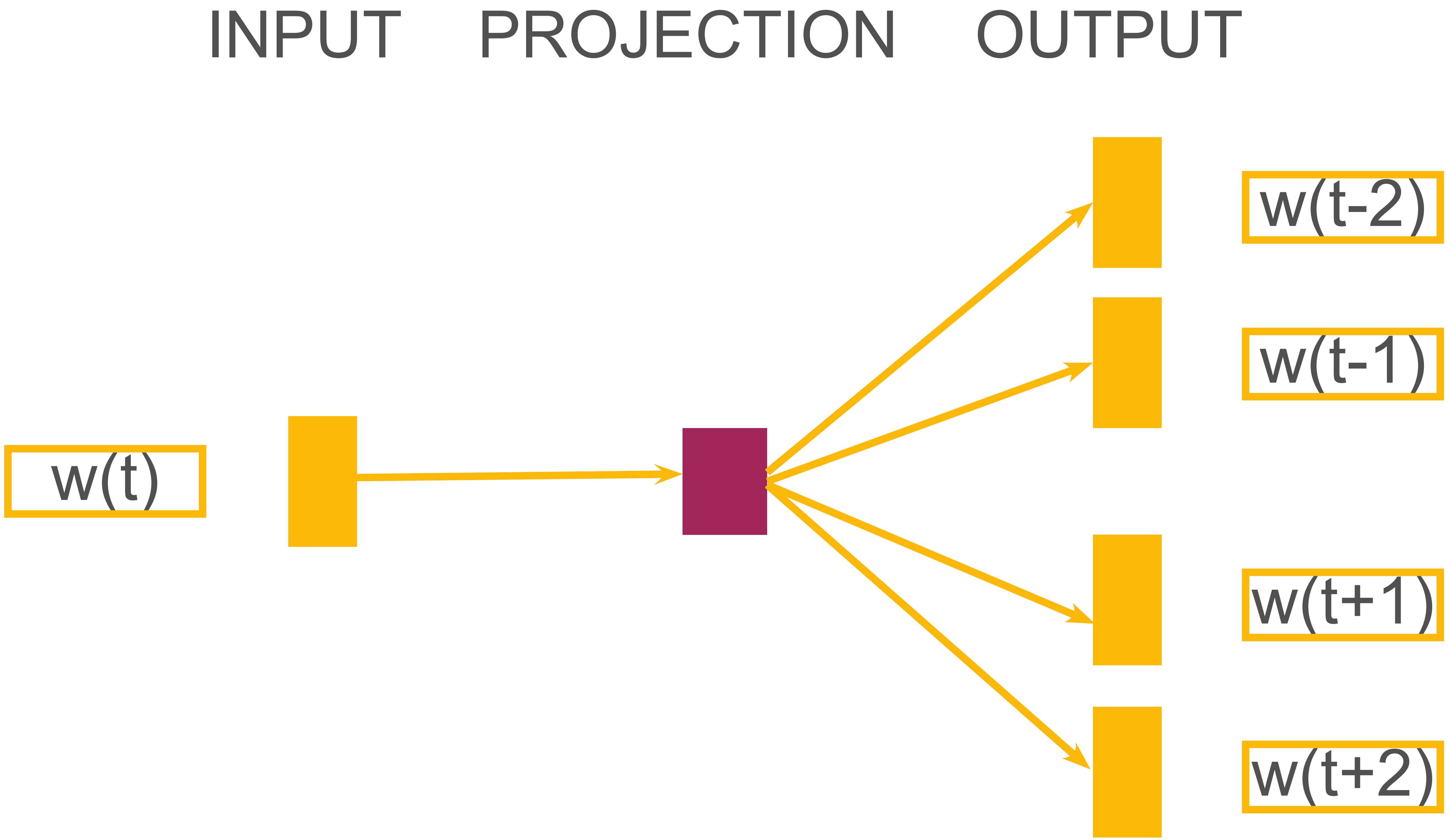


text analyzing in the sliding window
word prediction via supervised machine learning

Continuous bag-of-words (CBOW)



Skip gram



The principle of maximum likelihood

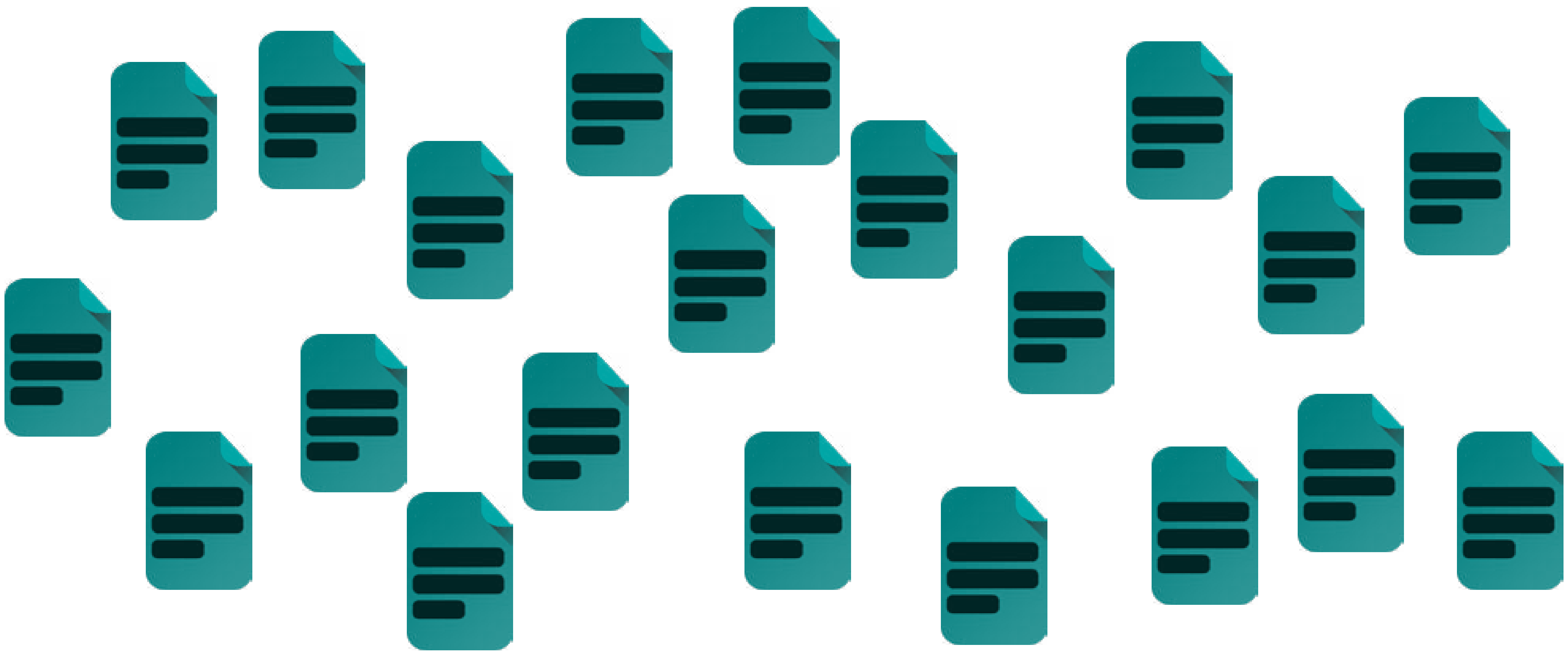
Each word ***w*** has two vectors:

- $\boldsymbol{v}_w$ - input
- $\boldsymbol{v'}_w$ - output

$$\operatorname{argmax} \frac{1}{T} \sum_{t=1}^{T} \sum_{-c \leq j \leq c,\, j \neq 0} \log p(w_{t+j} | w_t)$$

$$p(w_O | w_I) = \frac{\exp\left({v'_{w_O}}^T v_{w_I}\right)}{\sum_{w=1}^{W} \exp\left({v'_w}^T v_{w_I}\right)}$$

- Take a **large** corpus of documents (wikipedia, news articles)
- Fit word2vec model



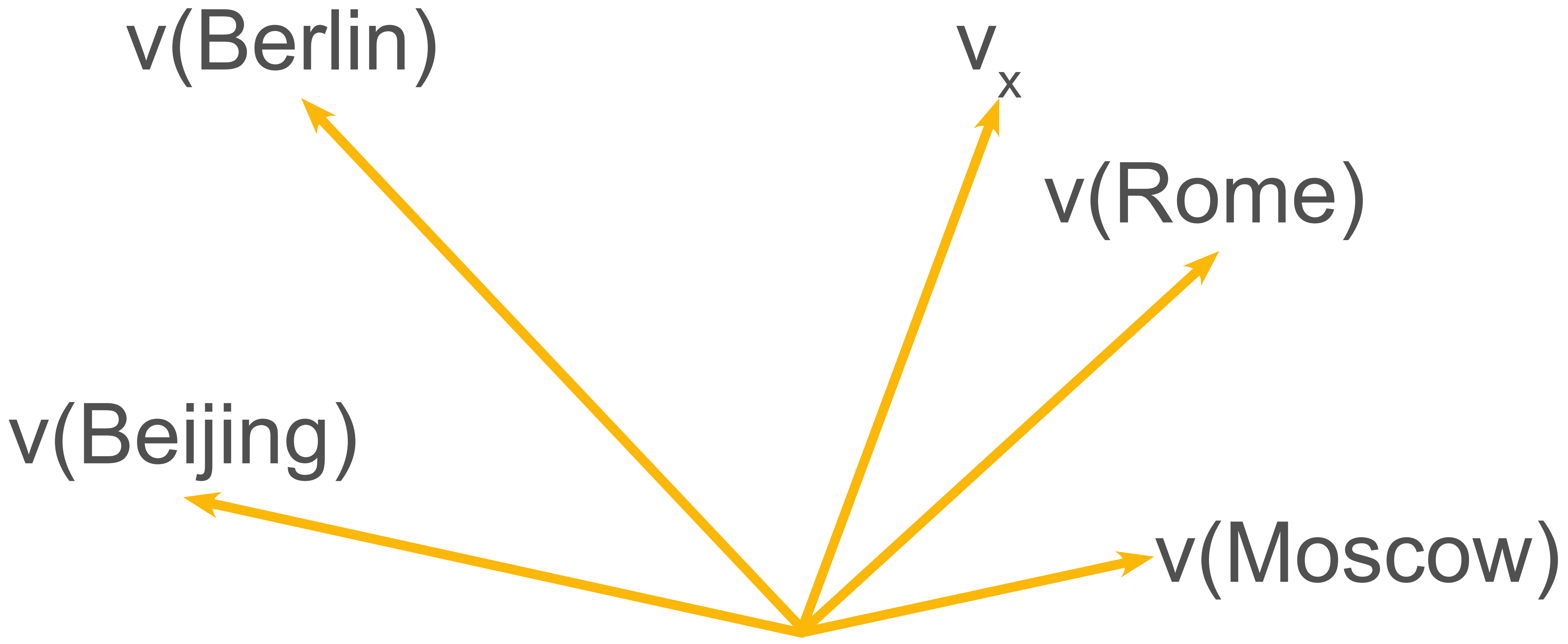
Question answering with Word2Vec

France : Paris,   Italy : ?

$$v_x = v(Paris) - v(France) + v(Italy)$$

$$v_x = v(Paris) - v(France) + v(Italy)$$

$$v(Rome) = \mathrm{argmax}_v \; (\cos(v_x, v))$$


v(Berlin)
v_x
v(Rome)
v(Beijing)
v(Moscow)

Type of relationship	Word Pair 1		Word Pair 2	
Common capital city	Athens	Greece	Oslo	Norway
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe
Currency	Angola	kwanza	Iran	rial
City-in-state	Chicago	Illinois	Stockton	California
Man-woman	brother	sister	grandson	granddaughter
Adjective to adverb	apparent	apparently	rapid	rapidly
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Superlative	easy	easiest	lucky	luckiest
Present participle	think	thinking	read	reading
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars
Plural verbs	work	works	speak	speaks

Word2vec answer

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Word2Vec answers

Application of word2vec:

- ► Text classification
- ► Machine translation
- ► Image captioning

Summary

- **Word2Vec** algorithm learns vector representations of words
- These vectors capture the “meanings” of words
- Similar words have similar vectors
- One can perform algebraic manipulations with these vectors