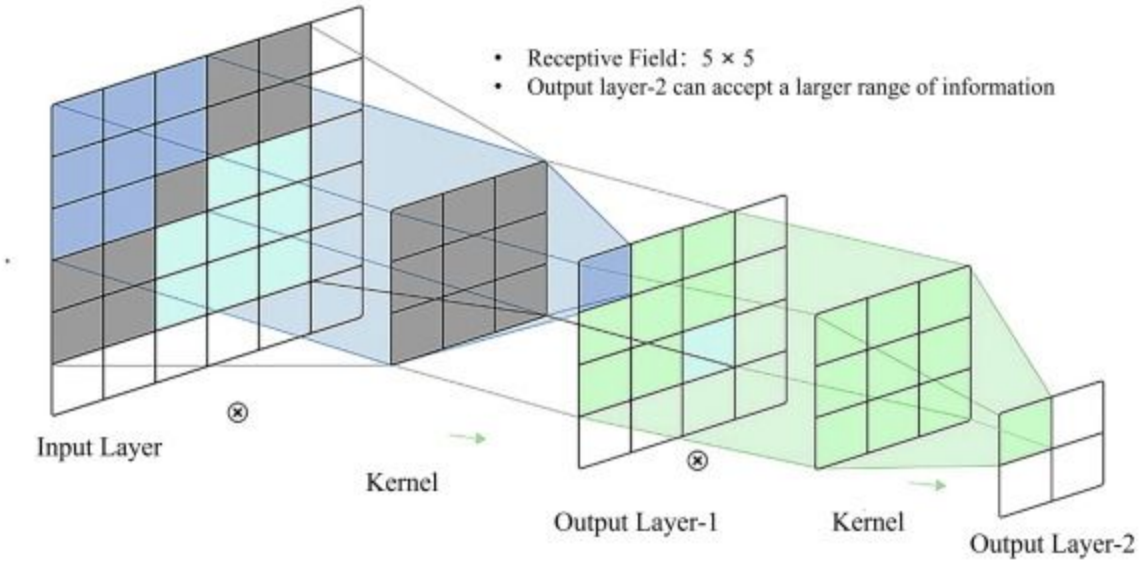
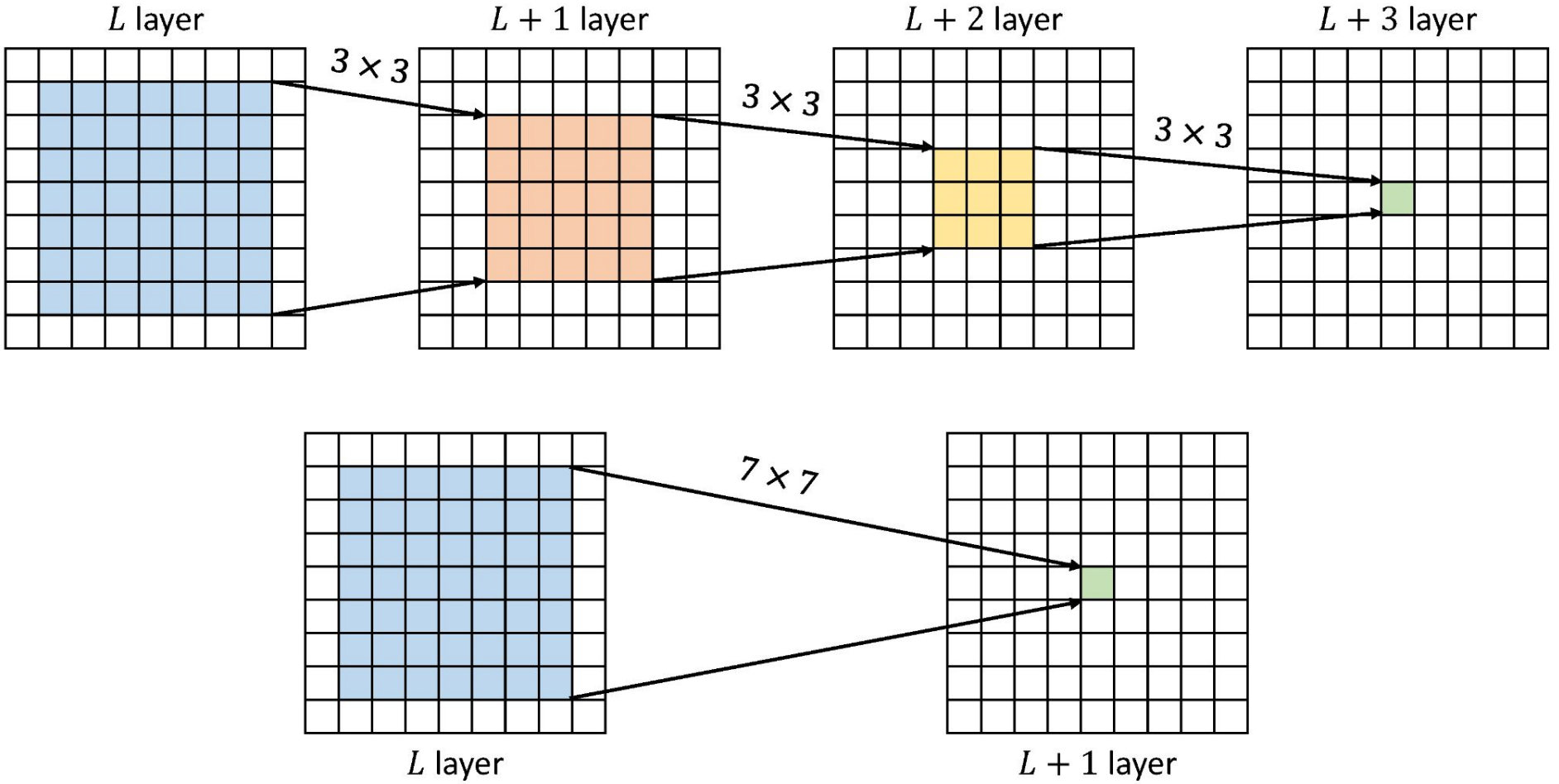
Embeddings

Марк Блуменау, магистратура ИИ

Повторка 1: Receptive field



[Receptive Field - an overview | ScienceDirect Topics]



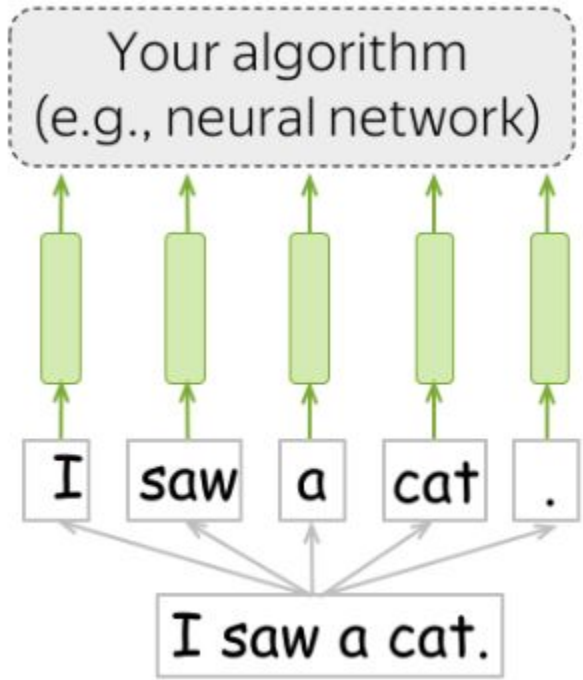
[Multi-Scale Receptive Fields Convolutional Network for Action Recognition]

Что вы уже знаете из ML

1) Токенизация
2) Bag of Words
3) TF-IDF
4) Стемминг
5) Лемматизация

Если не знаете – пересмотрите записи :)

Что вы узнаете сегодня



Any algorithm for solving a task

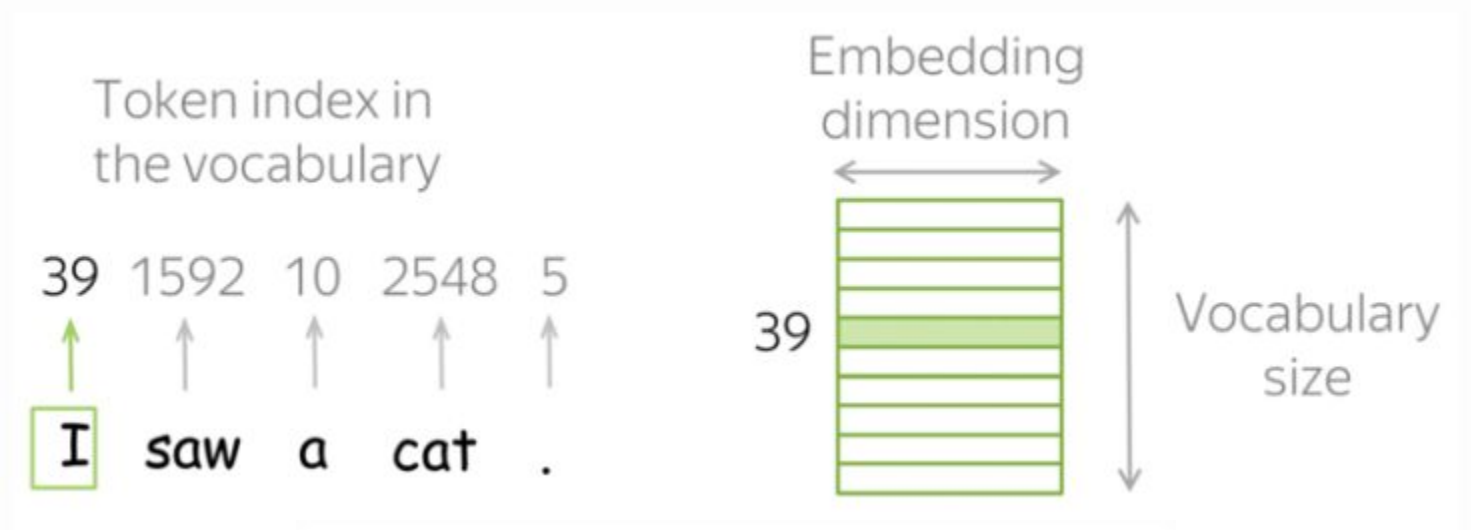
Word representation - vector
(input for your model/algorithm)

Sequence of tokens

Text (your input)

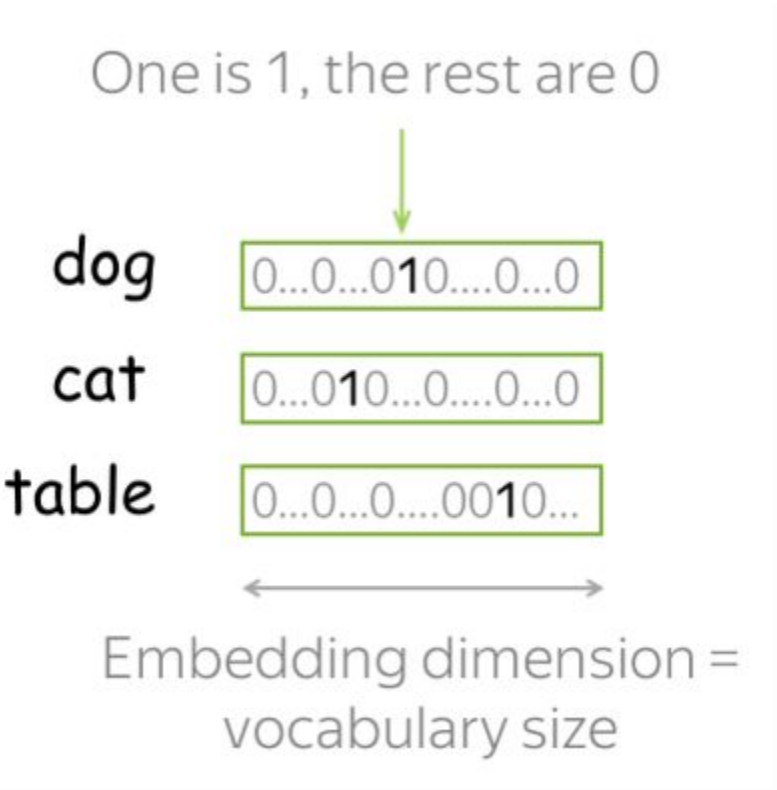
https://lena-voita.github.io/nlp_course/word_embeddings.html

Токенизация (напоминание)



https://lena-voita.github.io/nlp_course/word_embeddings.html

One hot или как не надо



https://lena-voita.github.io/nlp_course/word_embeddings.html

Что такое эмбеддинг?

Now look how this word is used in different contexts:

A bottle of tezgüino is on the table.

Everyone likes tezgüino.

Tezgüino makes you drunk.

We make tezgüino out of corn.

Can you understand what tezgüino means ?

https://lena-voita.github.io/nlp_course/word_embeddings.html

Восстановим слово по контексту

A bottle of tezgüino is on the table.

Everyone likes tezgüino.

Tezgüino makes you drunk.

We make tezgüino out of corn.

Tezgüino is a kind of alcoholic beverage made from corn.

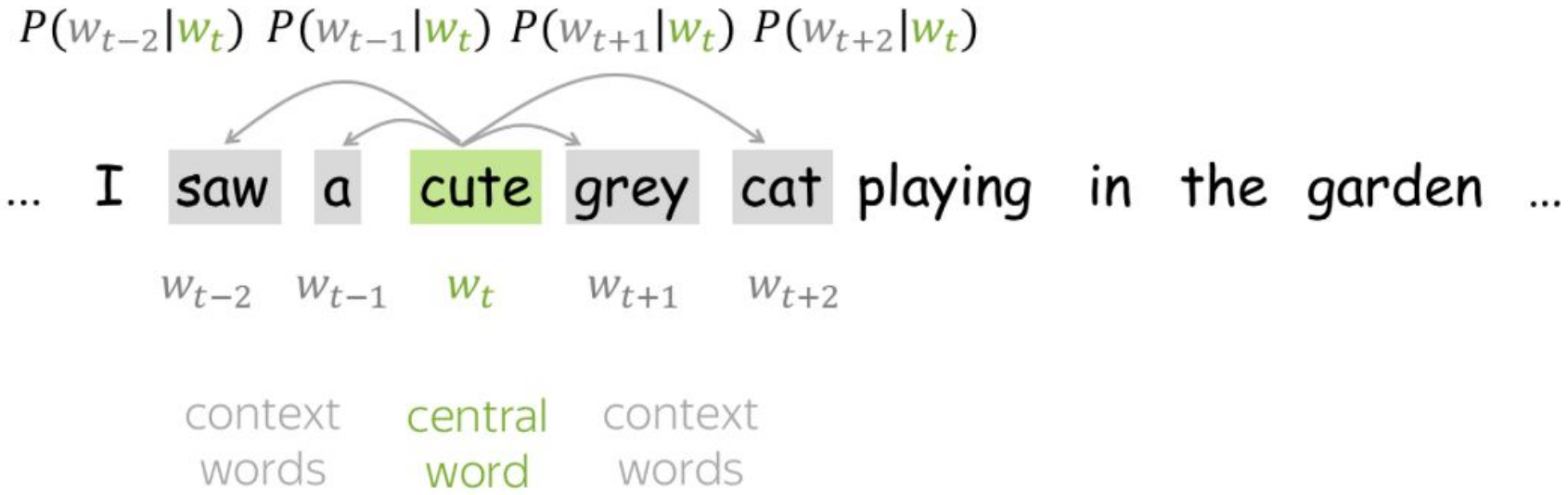
With context, you can understand the meaning!

https://lena-voita.github.io/nlp_course/word_embeddings.html

Запустим ваши нейронки в реалтайме

watch

Word2Vec



https://lena-voita.github.io/nlp_course/word_embeddings.html

Word2Vec

$$\text{Loss} = J(\theta) = -\frac{1}{T}\log L(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{-m\le j\le m,\\ j\neq 0}} \log P(w_{t+j}|w_t,\theta)$$

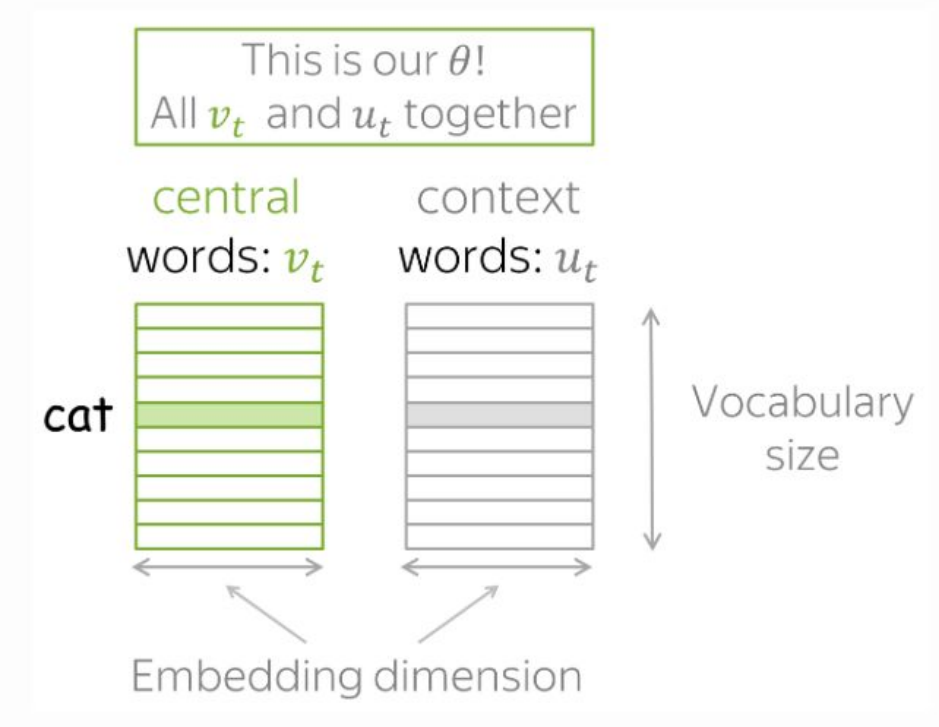
agrees with our plan above ⟼ go over text

with a sliding window

compute probability of the context word given the central

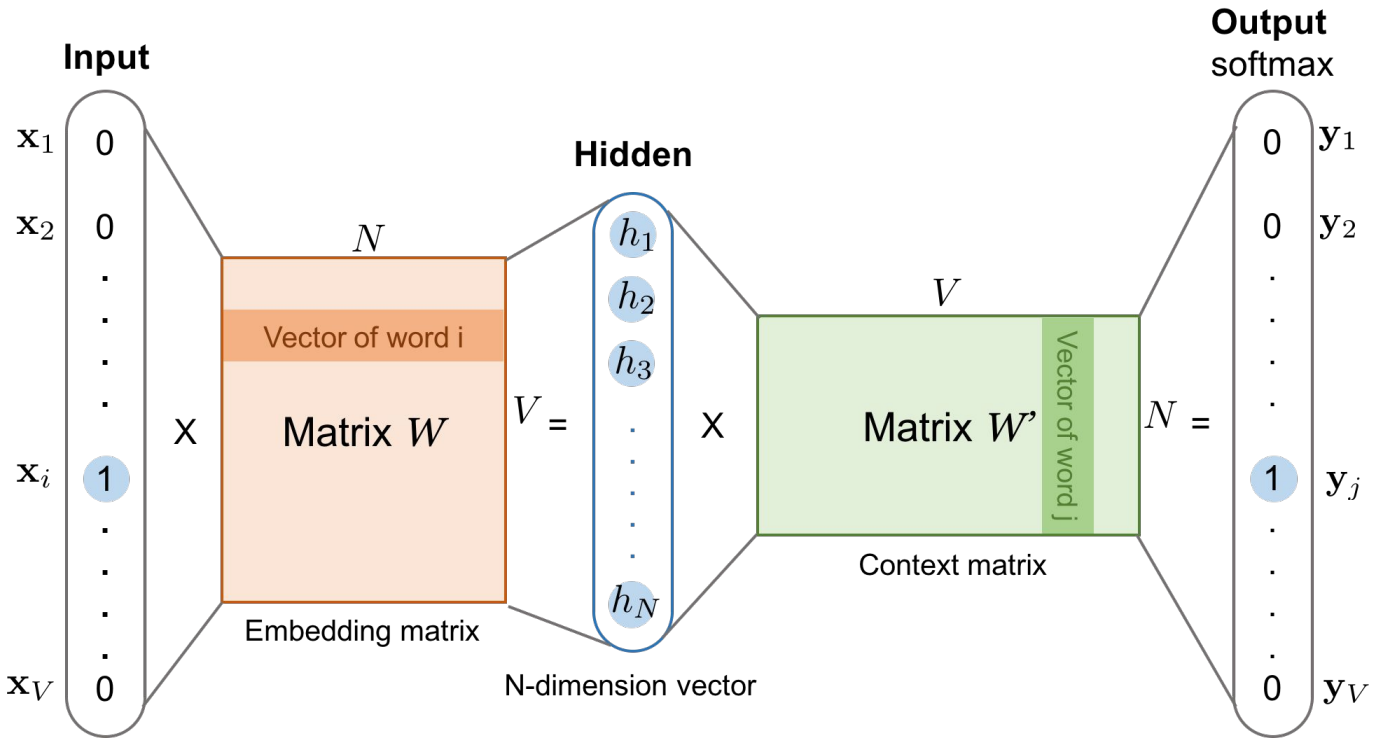
https://lena-voita.github.io/nlp_course/word_embeddings.html

Word2Vec



https://lena-voita.github.io/nlp_course/word_embeddings.html

Word2Vec



https://lilianweng.github.io/posts/2017-10-15-word-embedding/

Word2Vec

$$P(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)}$$

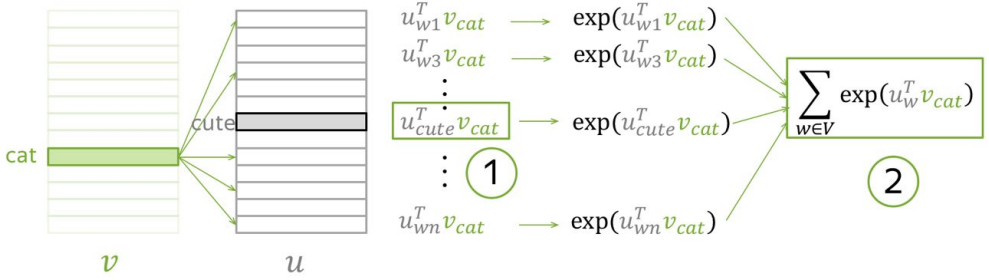
Dot product: measures similarity of o and c
Larger dot product = larger probability

Normalize over entire vocabulary
to get probability distribution

https://lena-voita.github.io/nlp_course/word_embeddings.html

Word2Vec

1. Take dot product of v_{cat} with **all** u 2. exp 3. sum all



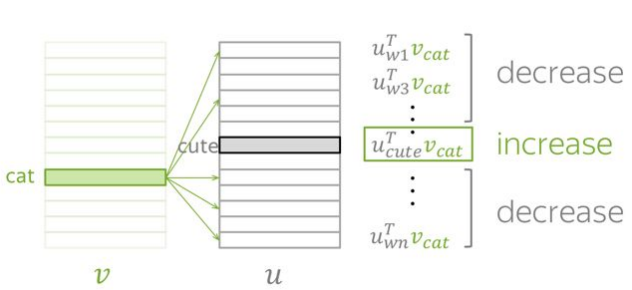
4. get loss (for this one step)

$$J_{t,j}(\theta) = \underbrace{-u_{cute}^T v_{cat}}_{1} + \underbrace{\log \sum_{w \in V} \exp(u_w^T v_{cat})}_{2}$$

5. evaluate the gradient, make an update

$$v_{cat} := v_{cat} - \alpha \frac{\partial J_{t,j}(\theta)}{\partial v_{cat}}$$

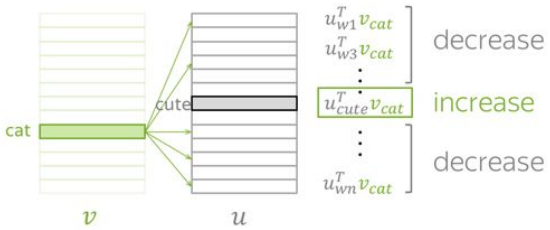
$$u_w := u_w - \alpha \frac{\partial J_{t,j}(\theta)}{\partial u_w} \; \forall w \in V$$



https://lena-voita.github.io/nlp_course/word_embeddings.html

Negative Sampling

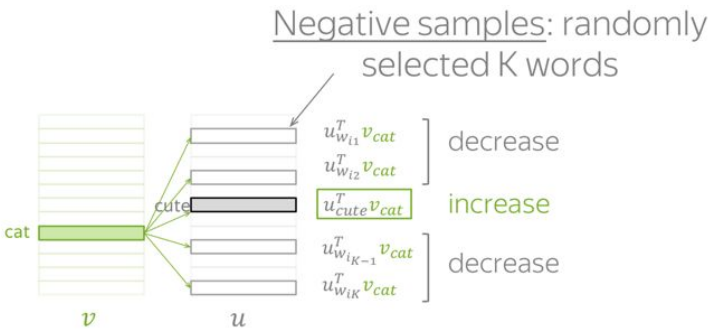
Dot product of v_{cat}:
- with u_{cute} - increase,
- with <u>all other</u> u - decrease



Parameters to be updated:
- v_{cat}
- u_w for all w in the vocabulary

|V| + 1 vectors

Dot product of v_{cat}:
- with u_{cute} - increase,
- with <u>a subset of other</u> u - decrease

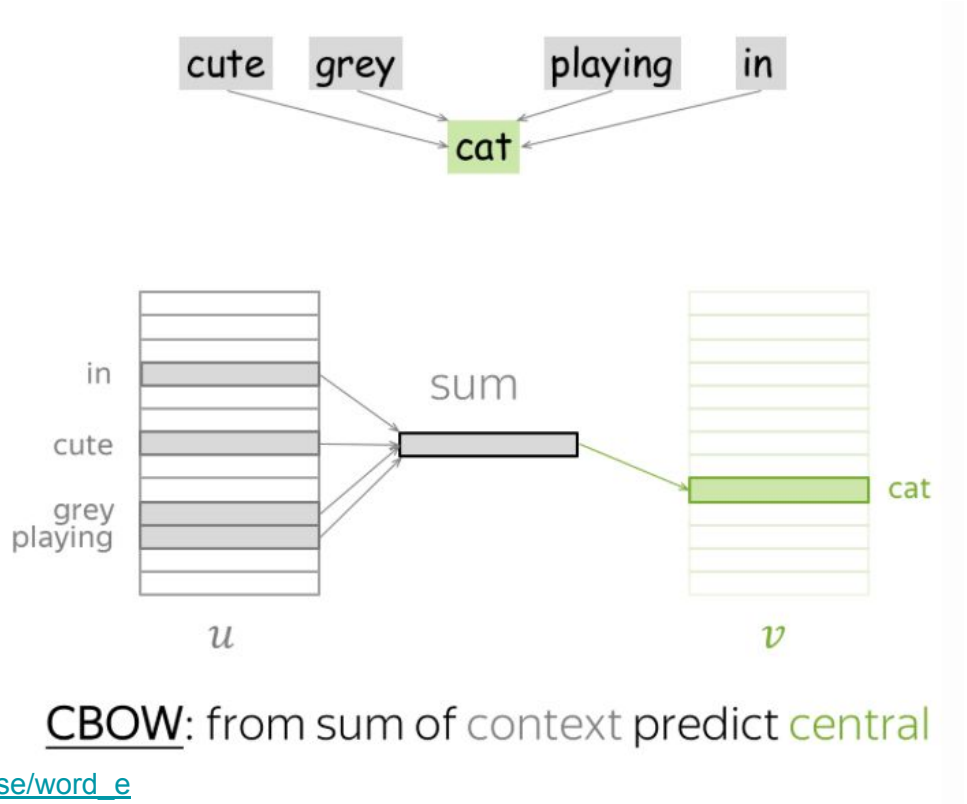


Parameters to be updated:
- v_{cat}
- u_{cute} and u_w for w in K negative examples

K + 2 vectors

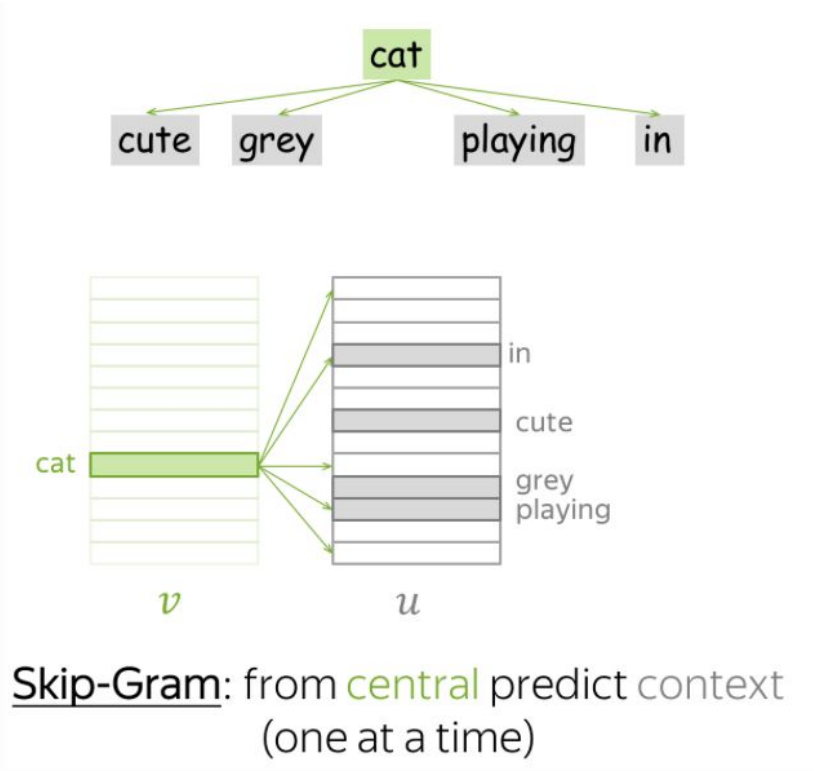
https://lena-voita.github.io/nlp_course/word_embeddings.html

Continuous Bag of Words aka CBOW



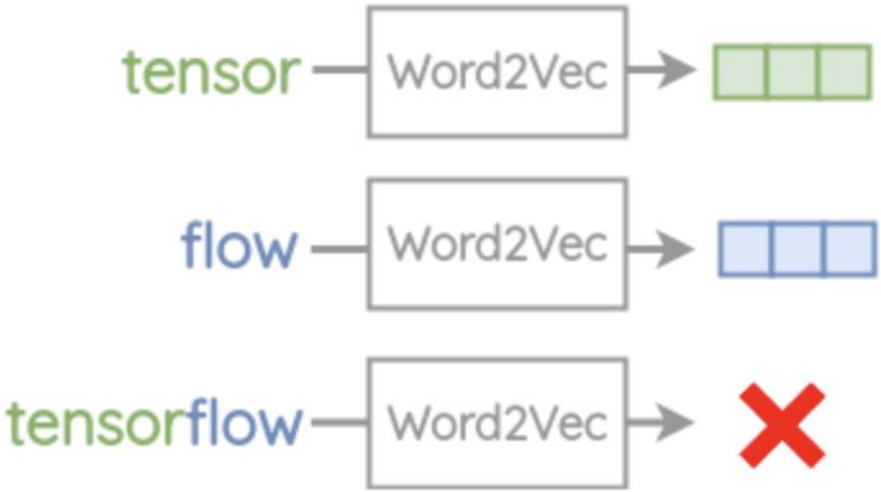
https://lena-voita.github.io/nlp_course/word_embeddings.html

Skip-gram



https://lena-voita.github.io/nlp_course/word_embeddings.html

FastText

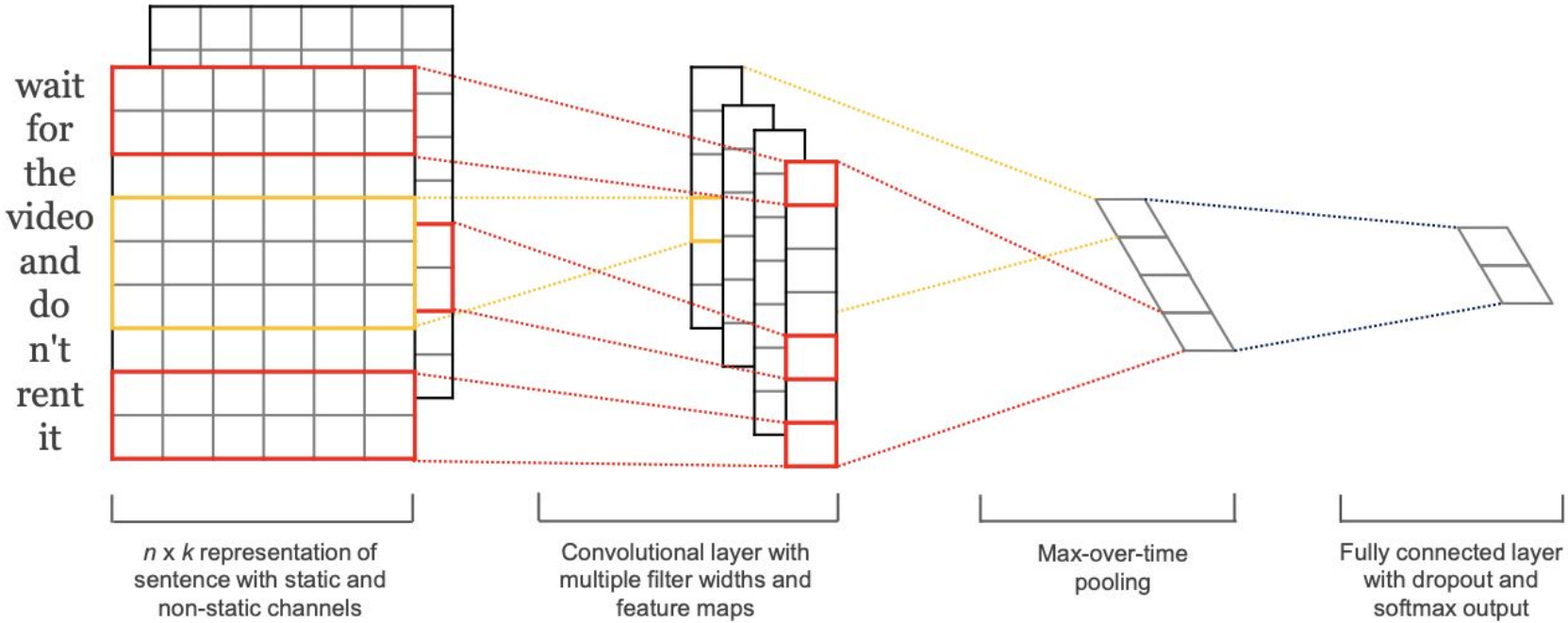


Shared radical

eat eats eaten eater eating

https://amitness.com/posts/fasttext-embeddings

Conv for text



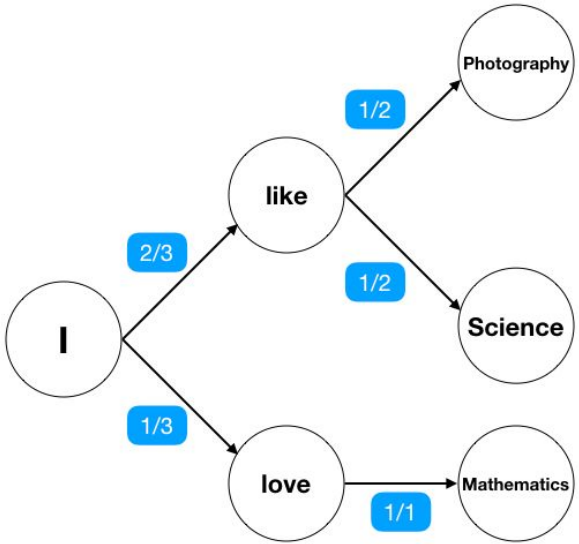
https://arxiv.org/abs/1408.5882

Почему так не всегда хорошо

1) Receptive field как бы не смотрит на весь текст
2) Как текст читают люди? А тут ведь не так…

Марковские модели

1) Наличие слова W обусловлено k словами до него
2) $p(w_1, \dots, w_n) = p(w_1)p(w_2|w_1) \dots p(w_n|w_{n-1}, \dots, w_{n-k})$
3) А теперь оценим вероятность aka как часто слово W встречается после k слов до него
4) Profit



[Next Word Prediction using Markov Model | by Ashwin M J | YML Innovation Lab | Medium]

Как сделать лучше?

RNN

Трансформер