Introduction

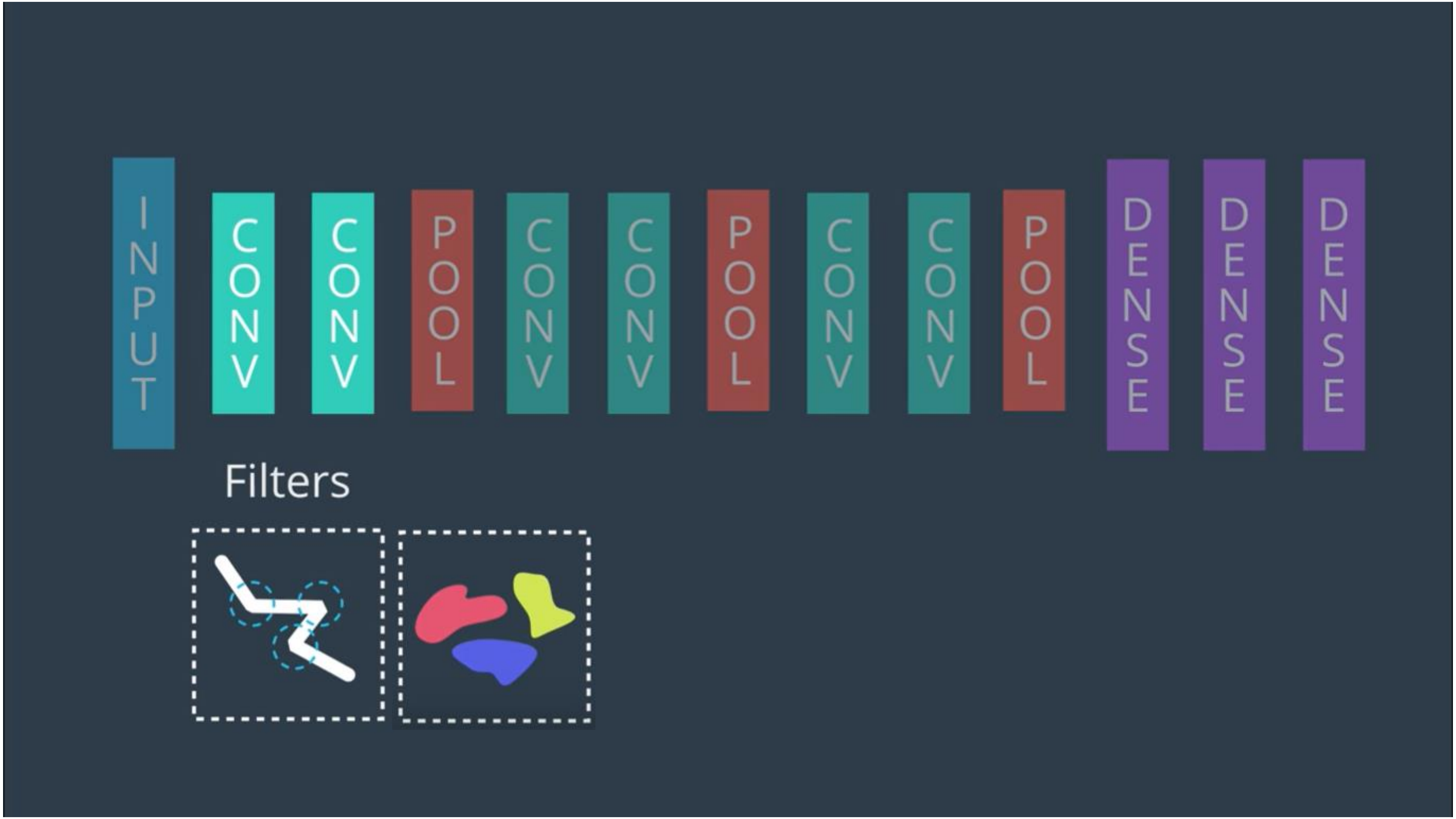
- 2018 was a breakthrough year in NLP. Transfer learning.

BERT

- BERT (Bidirectional Encoder Representations from Transformers). Google.
- You can either use these models to extract high quality language features from your text data,
or you can fine-tune these models on a specific task (classification, entity recognition, question answering, etc.)
with your own data to produce state of the art predictions.
- if you want to match customer questions or searches against already answered questions or well documented searches,
these representations will help you accuratley retrieve results matching the customer’s intent and contextual meaning,
even if there’s no keyword or phrase overlap.
- BERT offers an advantage over models like Word2Vec, because while each word has a fixed representation under
Word2Vec regardless of the context within which the word appears, BERT produces word representations that are dynamically
informed by the words around them. For example, given two sentences:

“The man was accused of robbing a bank.” “The man went fishing by the bank of the river.”

Word2Vec would produce the same word embedding for the word “bank” in both sentences, while under BERT the
word embedding for “bank” would be different for each sentence.



INPUT

CONV

CONV

POOL

CONV

CONV

POOL

CONV

CONV

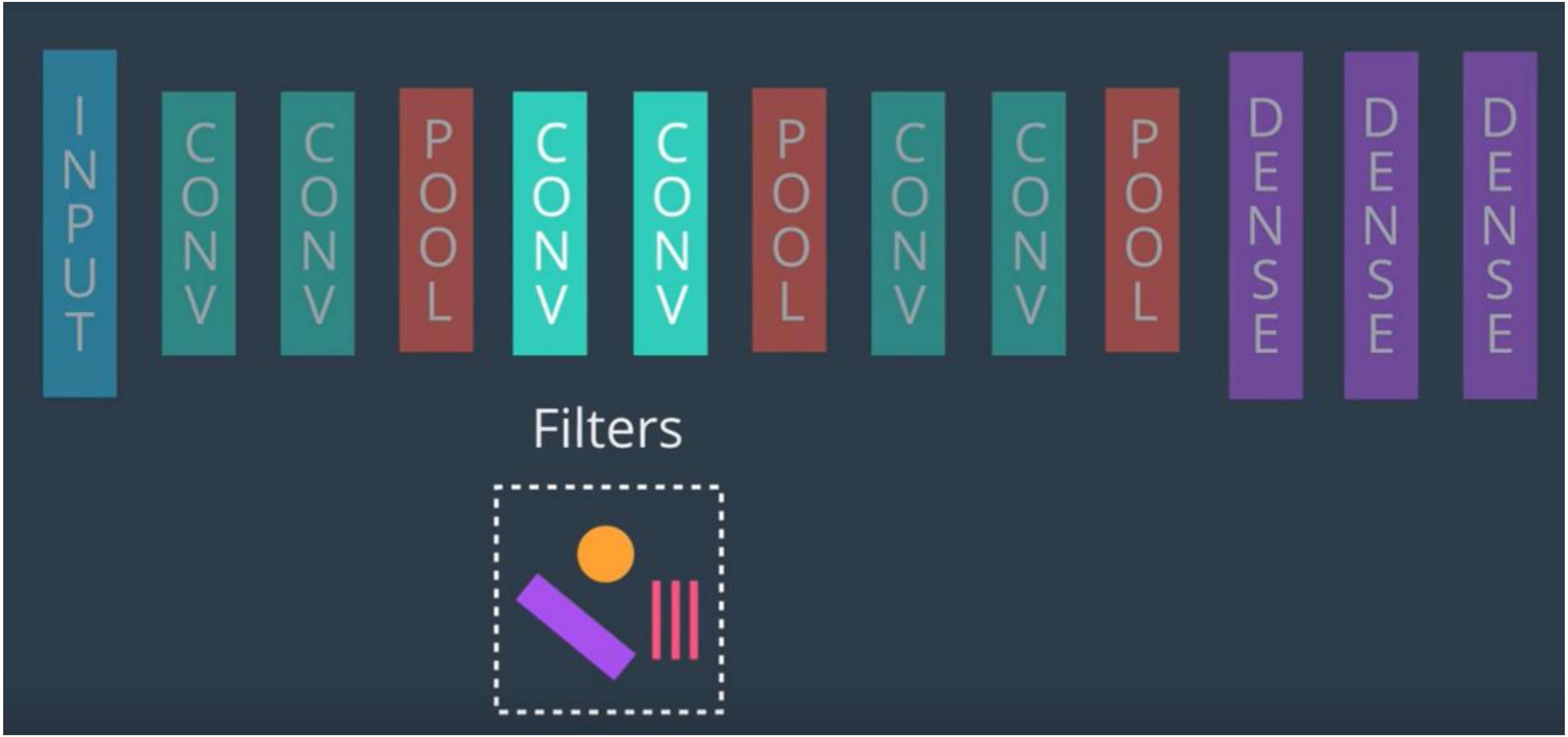
POOL

DENSE

DENSE

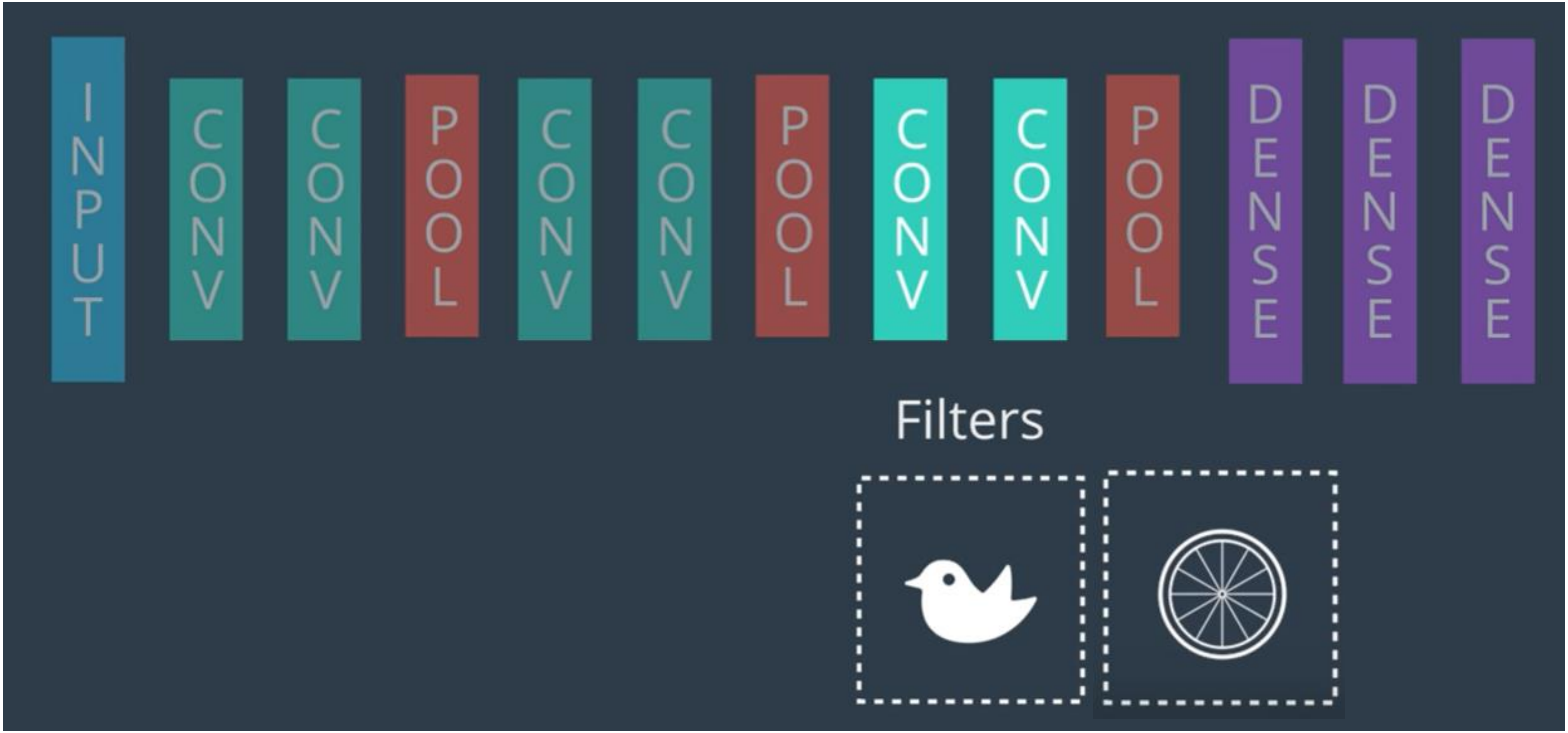
DENSE

Filters



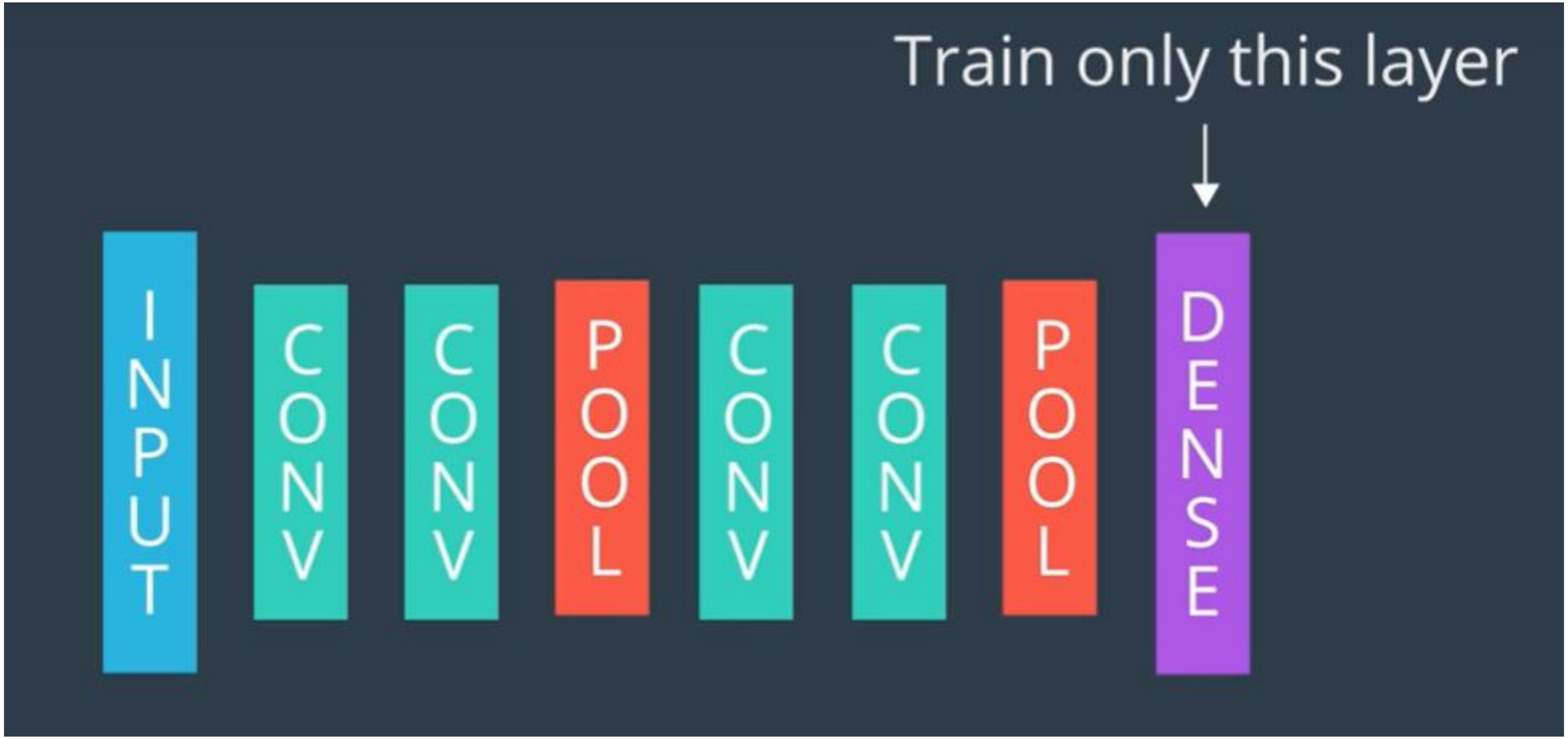
INPUT CONV CONV POOL CONV CONV POOL CONV CONV POOL DENSE DENSE DENSE

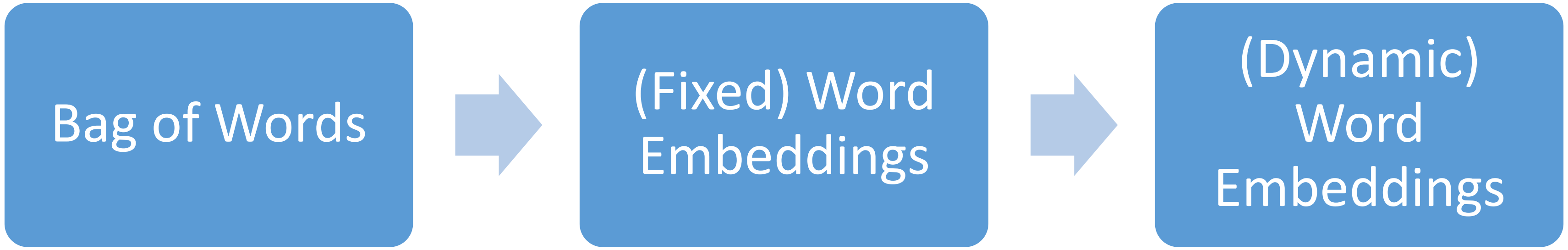
Filters



INPUT CONV CONV POOL CONV CONV POOL CONV CONV POOL DENSE DENSE DENSE

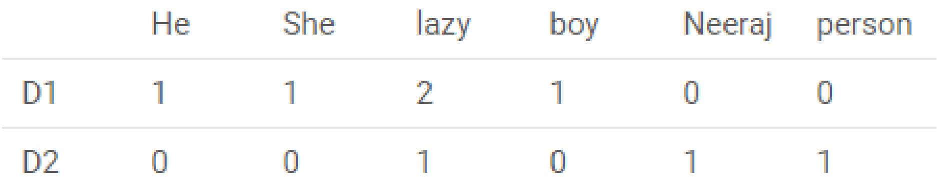
Filters


Train only this layer
INPUT
CONV
CONV
POOL
CONV
CONV
POOL
DENSE

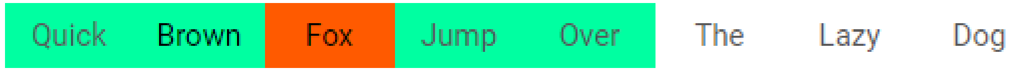


2 types of WE:

- Frequency based Embedding
  - Count Vector
  - TF-IDF Vector
  - Co-occurrence vector (Similar words tend to occur together and will have similar context for example)



	He	She	lazy	boy	Neeraj	person
D1	1	1	2	1	0	0
D2	0	0	1	0	1	1



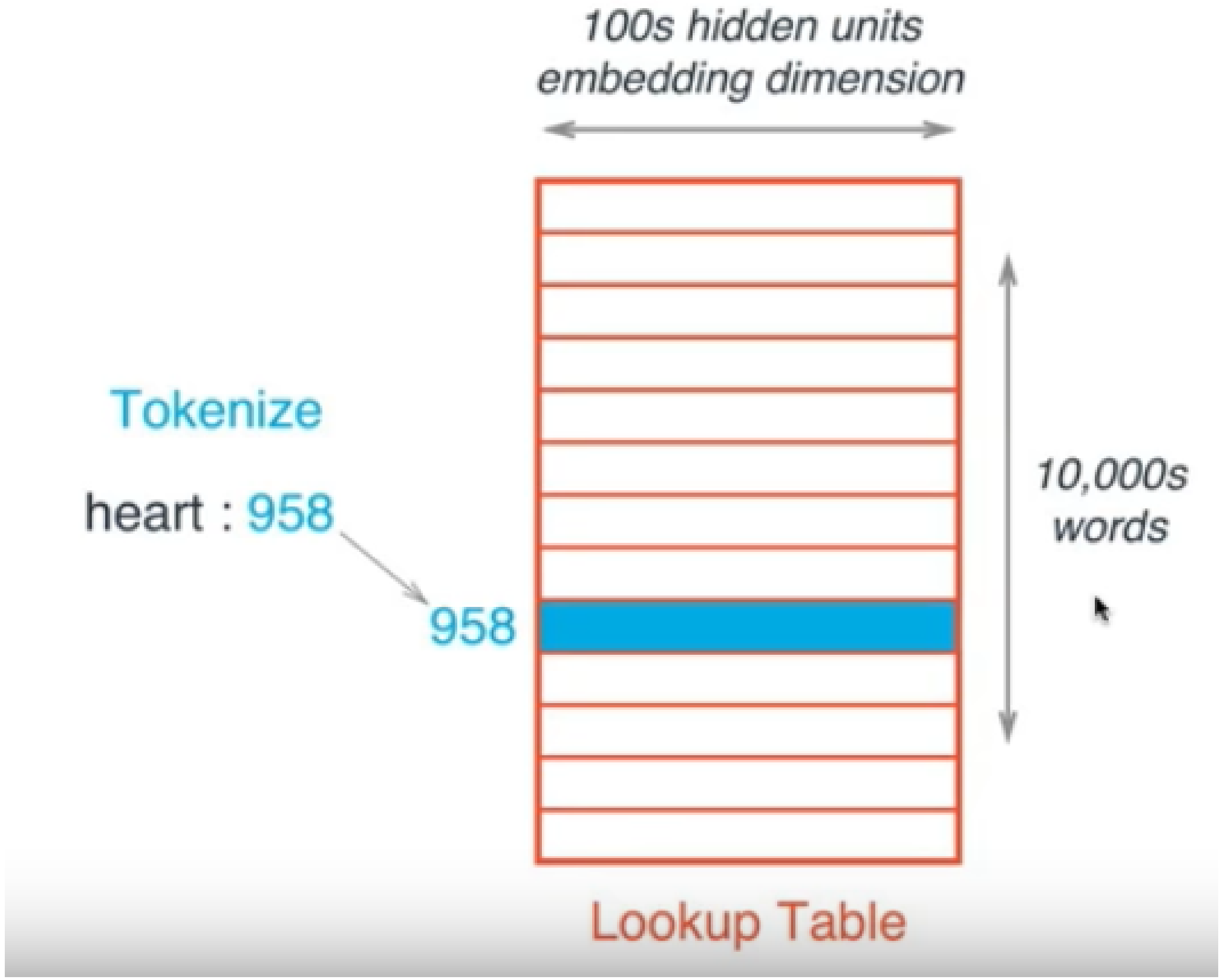
🟩 Quick	🟩 Brown	🟥 Fox	🟩 Jump	🟩 Over	The	Lazy	Dog

The green words are a 2 (around) context window for the word ‘Fox’ and for calculating the co-occurrence only these words will be counted.

- Prediction based Embedding

- Prediction based Embedding

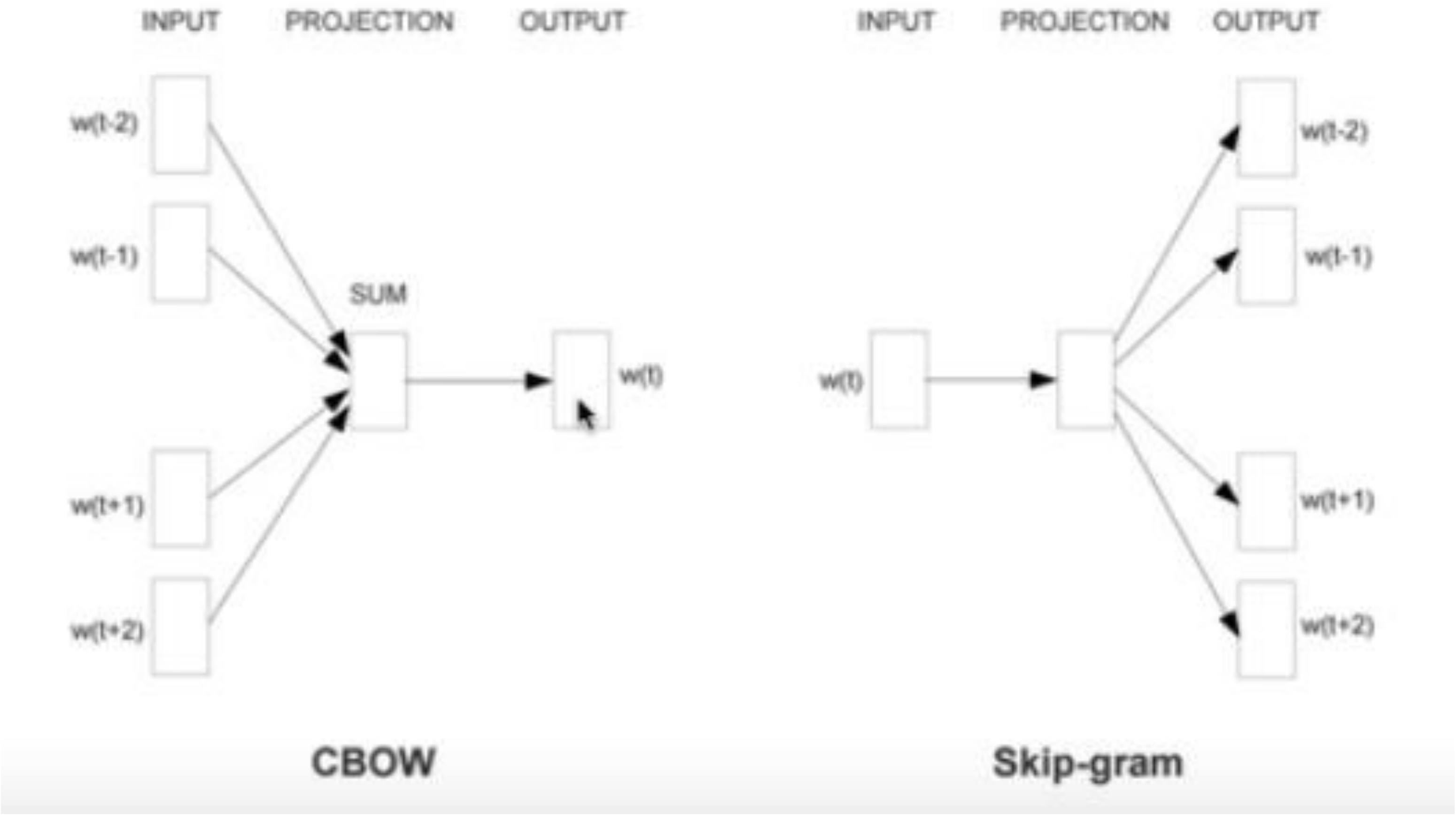
Not only for words. Whenever you have a massive number of classes



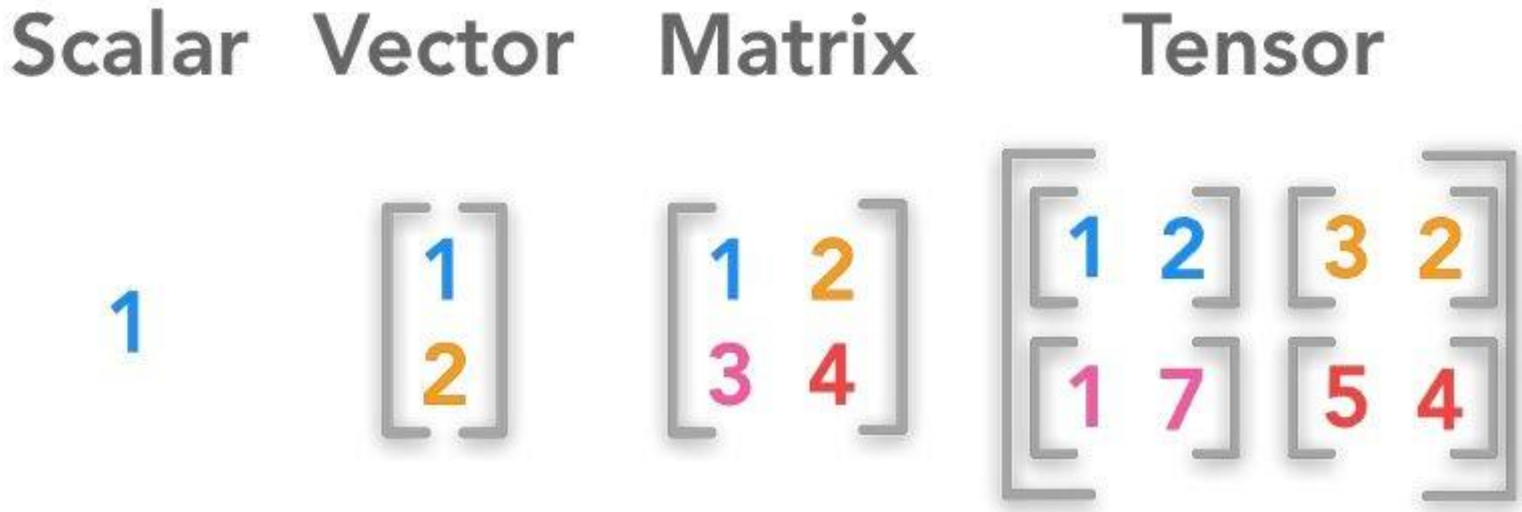
- Prediction based Embedding

Word2vec

- CBOW
  - Input is the context around your word
- Skip-gram
  - The other way round



- Analyzing BERT
  - Understand Tokenization in BERT
  - Run BERT & understand the output
  - Word & sentence vectors
  - Check similarity between vectors



buff.ly/2FSYAQK

Bibliografía:

https://www.analyticsvidhya.com/blog/2017/06/word-embeddings-count-word2veec/?utm_source=twitter.com

https://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial/