

**There are various approaches for converting
text into vector**

Label
Encoding

One Hot
Encoding

Bag Of Words
Bag of n-grams

TF-IDF

Word
Embeddings



Limitations of BOW and TF-IDF



add, auto, and, assistance, need, I, immediately, help, ... urgent (100 k words)

I need help [0 0 0 0 1 1 0 1 ... 0] ← vector size=100k



add, auto, and, assistance, need, I, immediately, help, ... urgent (100 k words)

I need help

[0 0 0 0 1 1 0 1 ... 0] ← vector size=100k

Sparse Representation



It may consume too much memory & compute resources

add, auto, and, assistance, need, I, immediately, help, urgent

I need help [0 0 0 0 0 1 1 0 1 0]

I need assistance [0 0 0 1 1 1 0 0 0]



add, auto, and, assistance, need, I, immediately, help, urgent

I need help [0 0 0 0 0 1 1 0 1 0]

I need assistance [0 0 0 1 1 1 0 0 0]



Doesn't capture meaning of words properly

Word embeddings try to address these shortcoming



Word embeddings try to address these shortcoming

Similar words have
similar vectors



Word embeddings try to address these shortcoming

Similar words have
similar vectors



Word embeddings try to address these shortcoming

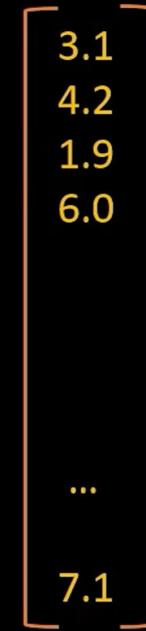
Similar words have similar vectors

Dimensions are low

good



great



Word embedding techniques

Word2vec

GloVe

fastText



Word embedding techniques

CBOW, Skip gram

Word2vec

GloVe

fastText



Word embedding techniques

CBOW, Skip gram

Based on transformer
architecture

Word2vec

GloVe

fastText

BERT

GPT



Word embedding techniques

CBOW, Skip gram

Word2vec

GloVe

fastText

Based on transformer
architecture

BERT

GPT

Based on LSTM

ELMo



Word2vec

King – man + woman

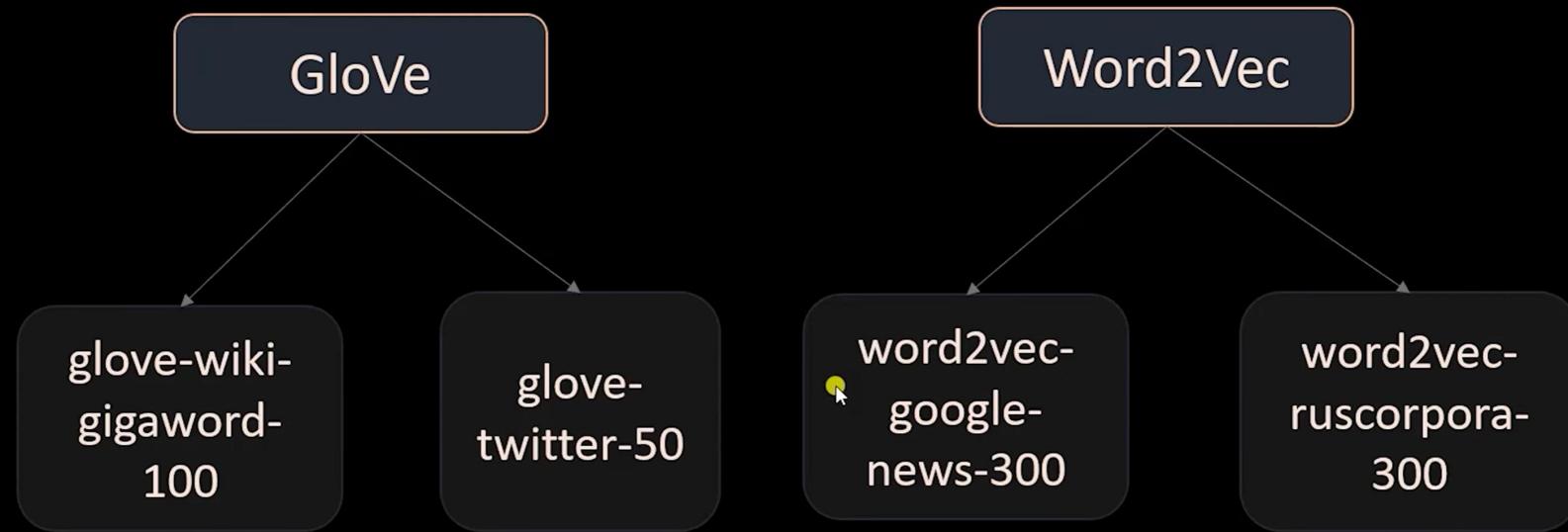


Word2vec

King – man + woman = Queen

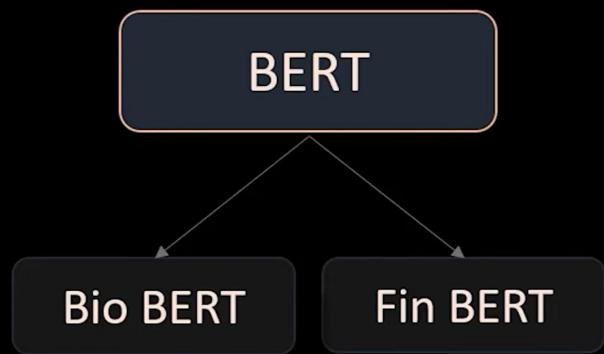


Variations based on dataset that they are trained on



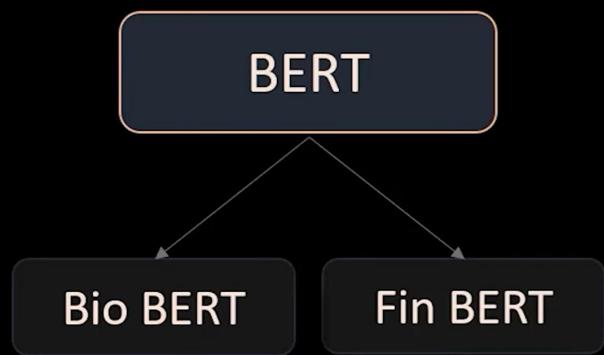
Variations of BERT models

Dataset or corpus on which
they are trained

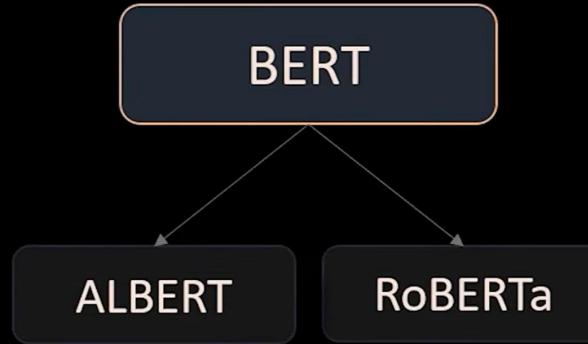


Variations of BERT models

Dataset or corpus on which
they are trained

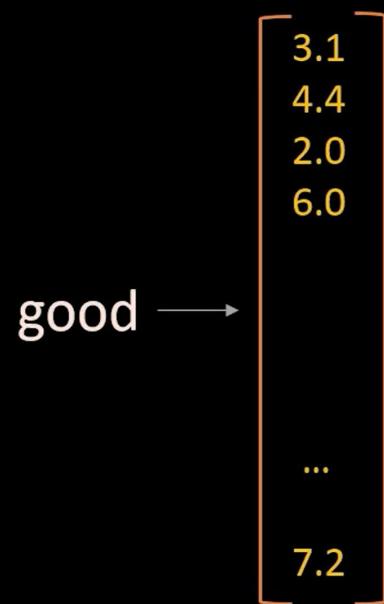


Parameter tuning



To summarize, these technique produce...

Word embedding



To summarize, these technique produce...

Word embedding



Sentence embedding

I am feeling good →



To summarize, these technique produce...

Word embedding



Sentence embedding



Doc embedding

