

Word Embeddings

$V = 10,000$ words

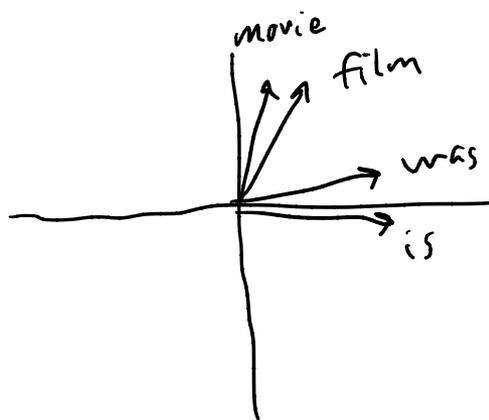
movie was good $\xrightarrow{\text{Bow}}$ $[0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ \dots \ 1 \ \dots]$
was good movie

$10,000$ -len vector \downarrow
 $= [0 \ 1 \ 0] + [0 \ 1 \ 0 \ \dots] + [0 \ 1 \ 0]$
movie was good

film is great $= [1 \ \dots \ 1 \ \dots \ 1]$
film is great

Orthogonal to "movie was good"

Word embeddings: low-dimensional representations of words
[50 - 300] capturing their similarity



How to learn embeddings

JR Firth 1957 Distributional hypothesis

"You shall know a word by the company it keeps"

I watched the movie

I watched the film

The film inspired me

The movie inspired me

I developed the film

in the darkroom

Brown Clusters, ...

Mikolov et al. 2013 "word2vec"

Predict each word's

Each word $w \rightarrow \vec{v}_w$ word vector
 $\rightarrow \vec{c}_w$ context vector

Context given
that word