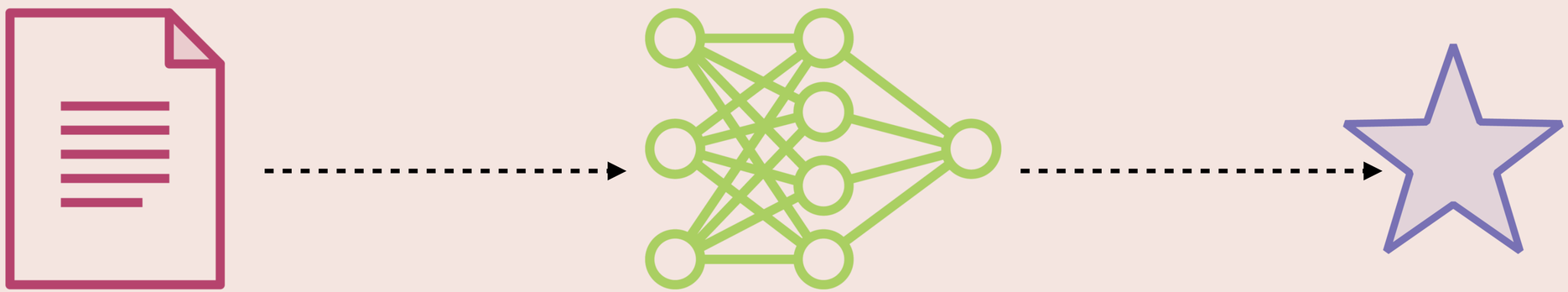
Training Word Representations

Axel Sirota
AI and Cloud Consultant

@AxelSirota

Reviews Dataset



Dataset

Model

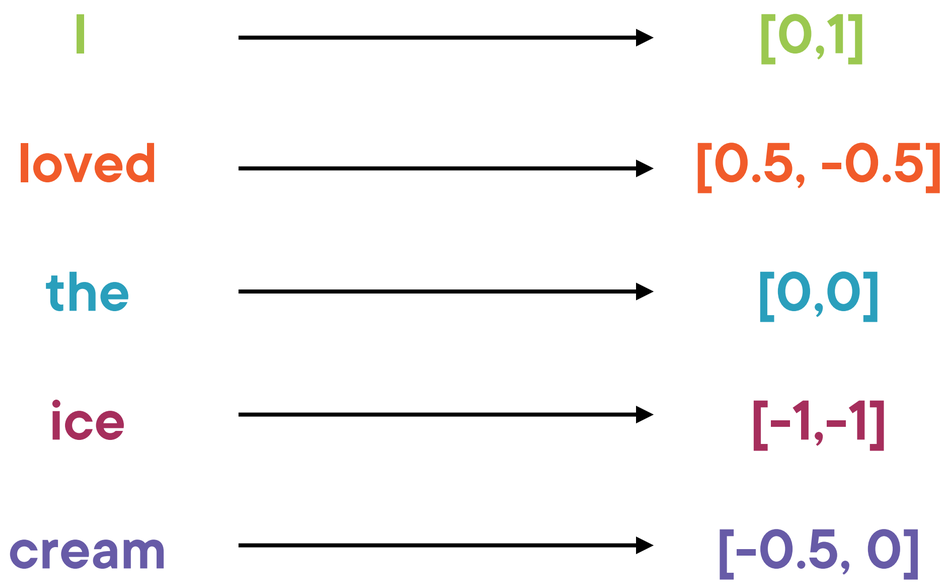
Stars

Training

I had a fabulous time ---------▸ Trained Model ---------▸ 5 stars

Evaluation

A Model Is Just a Function



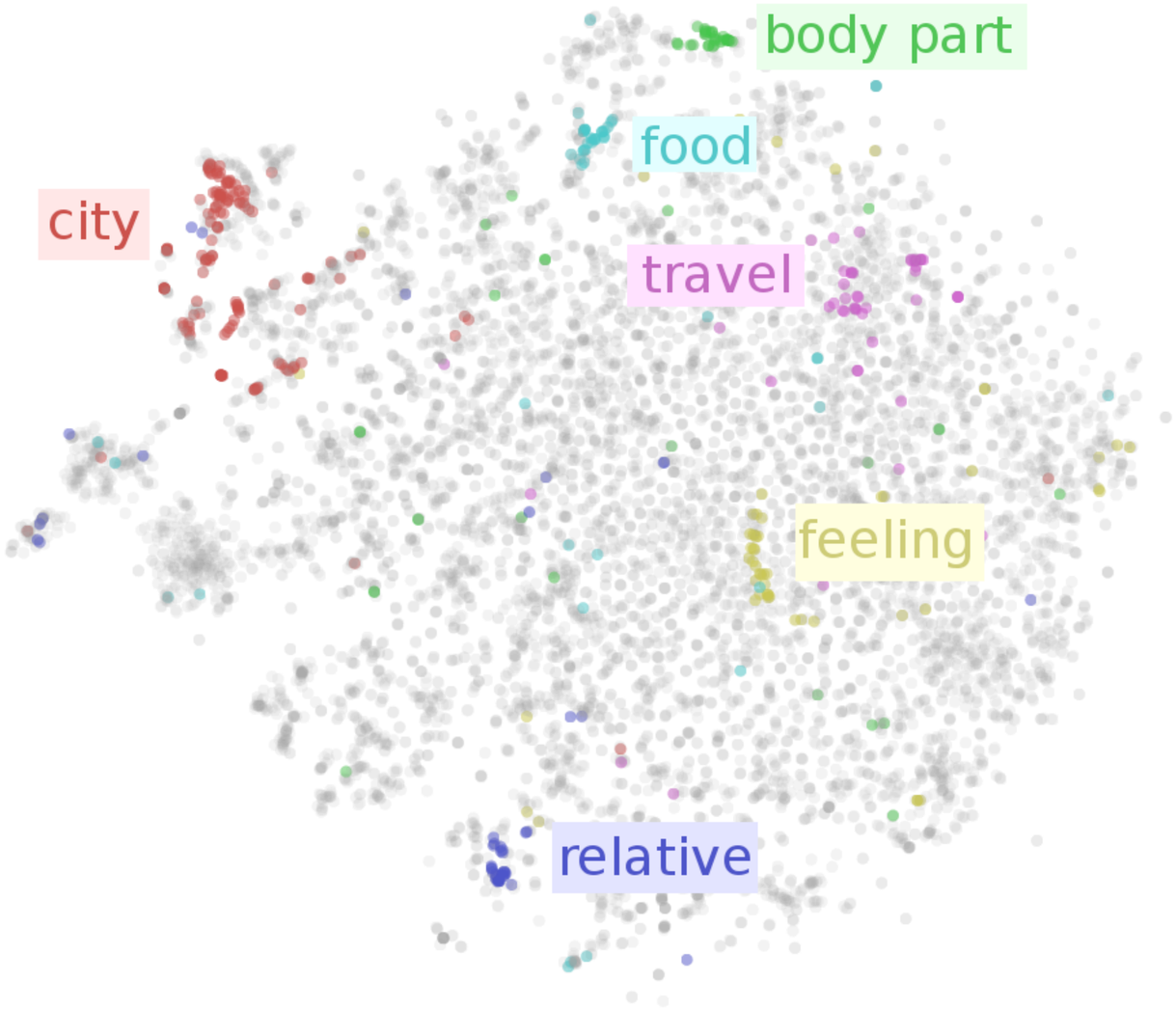
A Model Is Just a Function

“I loved the ice cream” →

0	**1**
0.5	-0.5
0	0
-1	-1
-0.5	0

This is known as an embedding

Representing Word Clusters



First Embedding: One Hot Encoding

Representing a Sentence

1. Construct vocabulary and assign different columns for presence

I loved the ice cream and loved the food

Vocabulary = {I, loved, the, ice, cream, and, food}

Representing a Sentence

I loved the ice cream and loved the food

1. Construct vocabulary and assign different columns for presence

2. Count

I	loved	the	ice	cream	and	food
1	0	0	0	0	0	0
0	**2**	0	0	0	0	0
0	0	**2**	0	0	0	0
0	0	0	**1**	0	0	0
0	0	0	0	**1**	0	0
0	0	0	0	0	**1**	0
0	0	0	0	0	0	**1**

Representing a Sentence

I loved the ice cream and loved the food

Final result: [1,2,2,1,1,1,1]

1. Construct vocabulary and assign different columns for presence

2. Count

[1,2,3]

3. Collapse into a vector

Demo

Represent the Yelp reviews dataset with One Hot Encoding

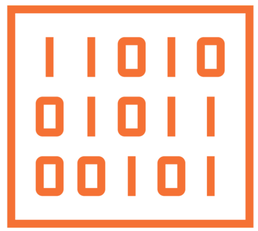
Demo

Predict Sentiment with OHE on Yelp dataset

Training Embeddings with Networks: CBOW and Skip-gram

NLP Exploded Quite Recently

Foundational paper: Efficient Estimation of Word Representations in Vector Space



We can learn embeddings with less than 500 dimensions (way less than OHE or other linguistic approaches)

Big development of embeddings: GloVe, FastText, Bort, Bert, GPT, etc...

Self-Construct a Supervised Learning Problem

Construct supervised learning problem

Train neural network

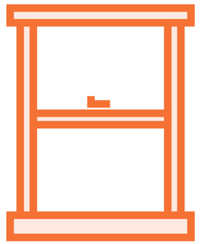
Extract first layer

Word2Vec Architectures

Continuous Bag of Words (CBOW)

Skip Gram

Defining CBOW

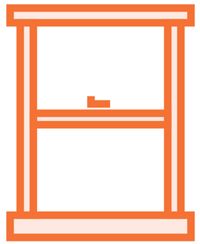


Define window size

CBOW problem

Defining CBOW

CBOW problem

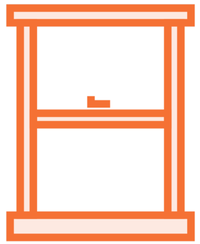


Define window size

Slide the window and take as input the remaining 4 words

Defining CBOW

CBOW problem



Define window size

Slide the window and take as input the remaining 4 words

[1,2,3]

3. Network predicts the middle word

Defining CBOW

I love ice cream so much I could eat it all day

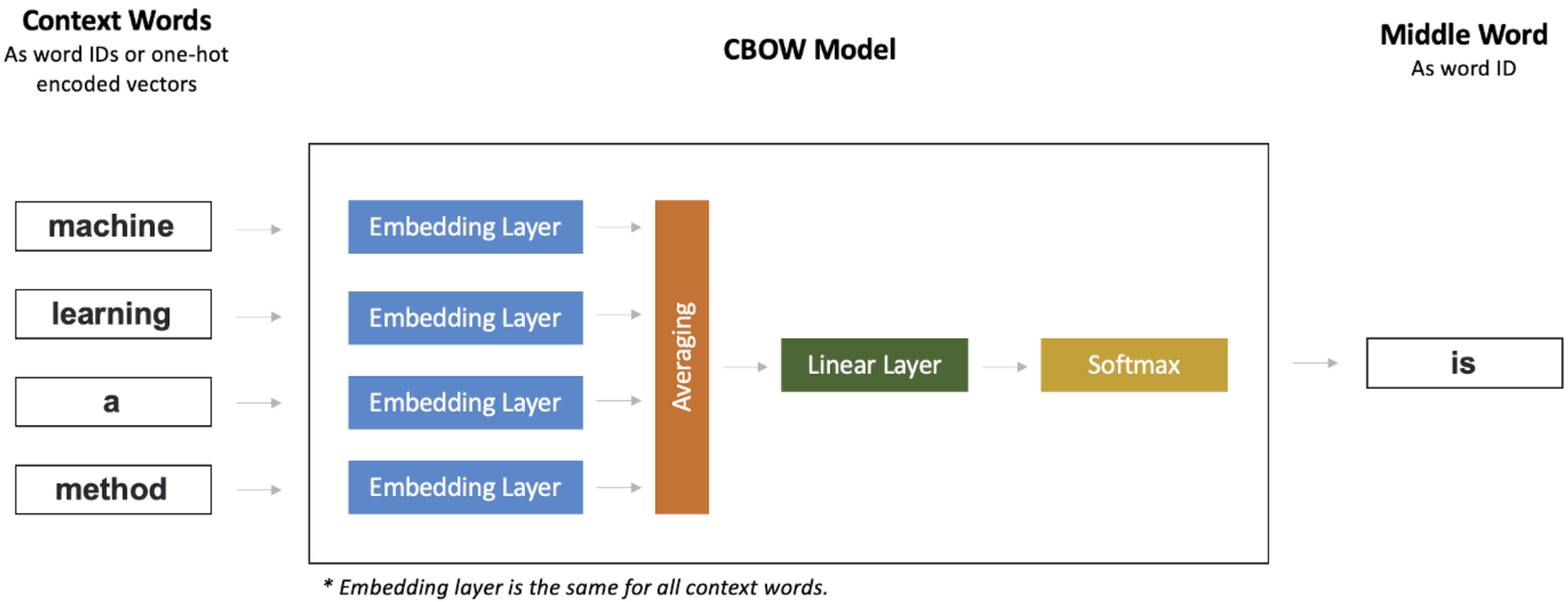
Input	Target
[“I”, “love”, “cream” ,”so”]	“ice”
[“love”, “ice”, “so” ,”much”]	“cream”
[“ice”, “cream”, “much”, “I”]	“so”

Defining CBOW

I love ice cream so much I could eat it all day

Input	Target
[2427,12, 546 ,853]	987
[12, 987, 853 ,431]	546
[987, 546, 431, 2427]	853

Defining CBOW



Source: Efficient Estimation of Word Representations in Vector Space

Demo

Train a CBOW model

Demo

Reanalyse sentiment with an existent embedding

Takeaways

An embedding is a mapping representation from text tokens into a numerical form

One hot encoding is the simplest embedding but it doesn't collate the information about word closeness

We can build word embeddings with deep learning

One can input OHE representation and learn the embedding layer into any task or one can use a pre-trained layer

Keys

Practice creating a Skip gram embedding

Try to ensure you understand at every step the different dimensions of the tensors and why they make sense

Practice creating a sentiment analyser out of the 300 dimensional glove vectors

Up Next:
Fine Tuning Word Representations