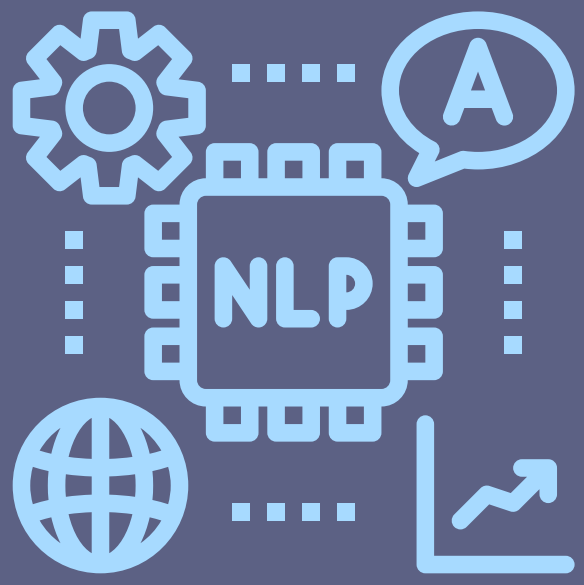


NATURAL LANGUAGE PROCESING

NLP is a field of AI that enables computers to understand, interpret, and generate human language.

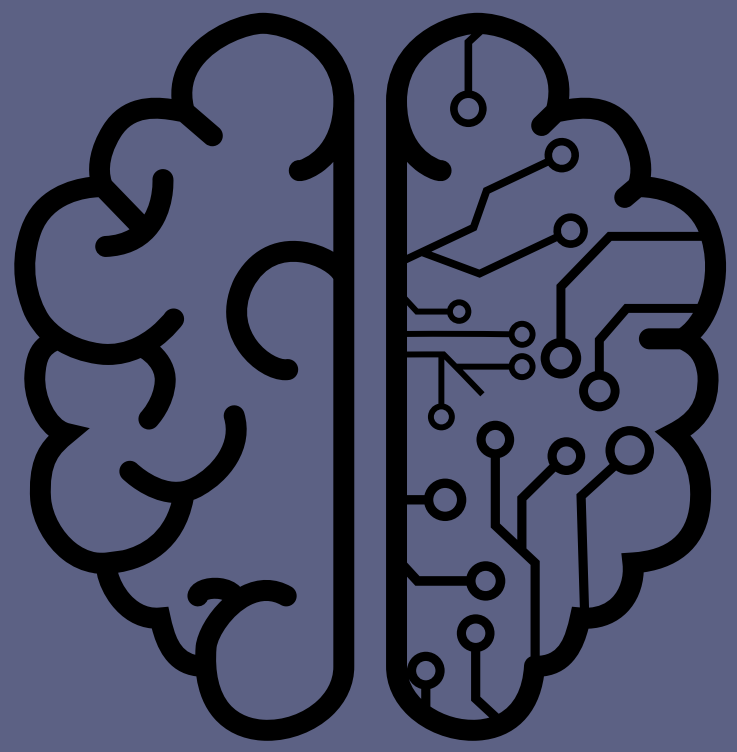
WORD2VEC

- **Continuous Bag of Words**
- **Skip Grams**

Aditya Goyal

Introduction

Word embeddings in NLP are real-valued vectors that capture the meanings of words, allowing similar words to be close in vector space.



- Developed by Google in 2013.
- Computes vector representations of words to capture semantic meaning.
- Uses deep learning and neural networks

Advantages:

- Low-Dimensional Vectors: Enables fast computation.
- Dense Vectors: Avoids sparsity.

Intuition

Let's take an example to understand word vectors.

- Words have vector representations based on features created by neural networks.
- This helps us grasp semantic relationships, despite deep learning's black-box nature.

Word	Gender	Wealth	Power	Weight	Speak
King	1	1	1	0.8	1
Queen	0	1	0.7	0.4	1
Man	1	0.3	0.2	0.6	1
Woman	0	0.3	0.2	0.5	1
Monkey	1	0	0	0.3	0

"Word2Vec assumes that words appearing in similar contexts have similar meanings and thus receive similar vector representations from the model."

“ King - Man + Woman = ? ”

"Using vector calculations, we can explore relationships, where a predicted vector is calculated internally to reveal semantic connections."

Feature	King	-	Man	+	Woman	Predicted Vector	Actual Vector (Queen)
Gender	1	-	1	+	0	0	0
Wealth	1	-	0.3	+	0.3	1	1
Power	1	-	0.2	+	0.2	1	0.7
Weight	0.8	-	0.6	+	0.5	0.7	0.4
Spark	1	-	1	+	1	1	1

The predicted vector is often very close to the actual vector of 'Queen'

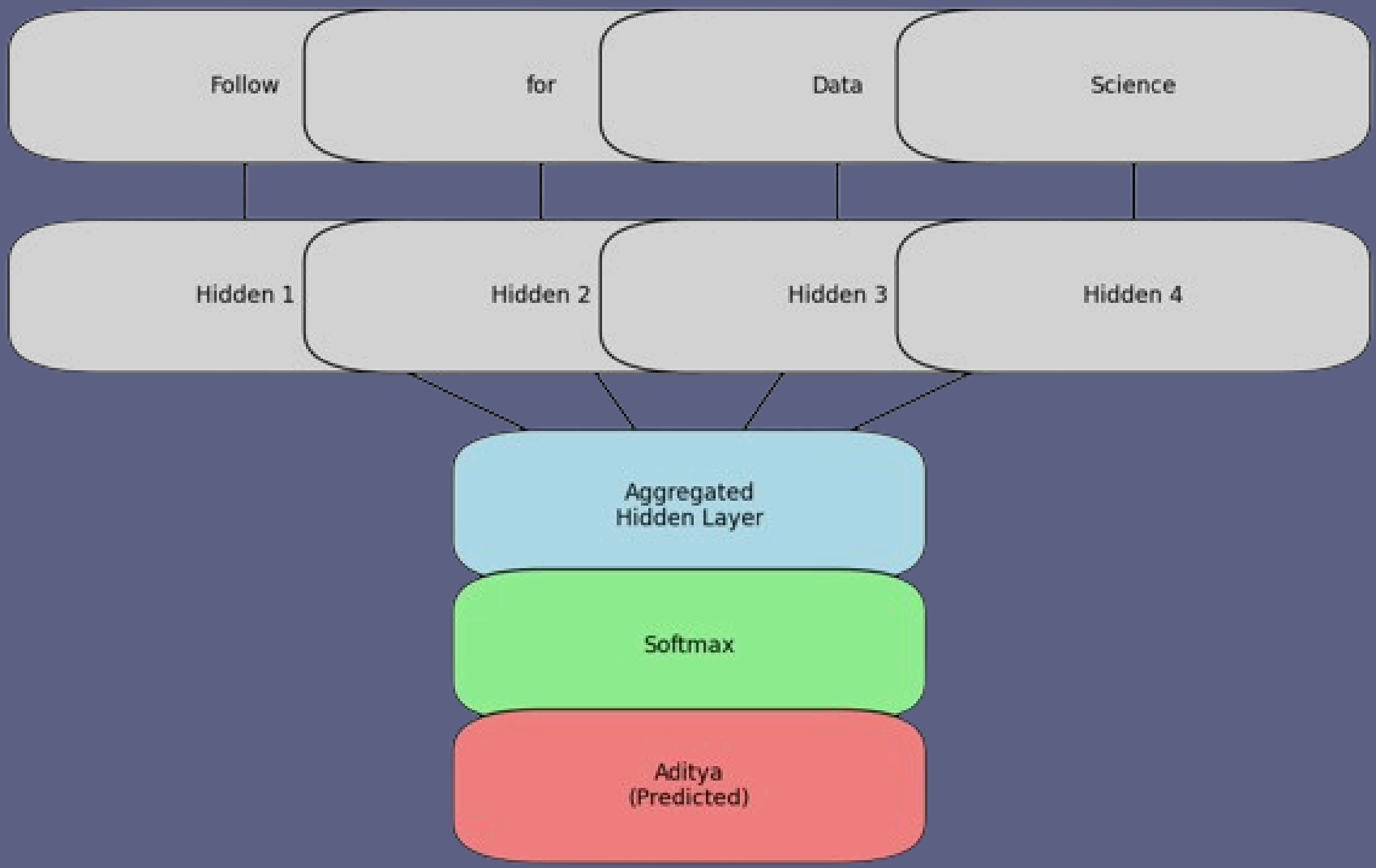
Vector Representation

To convert words to vectors, we introduce and solve a fake problem, obtaining the desired vector as a byproduct.

Sentence	Target Word	Context Words
Follow Aditya for Data Science	Aditya	Follow, for, Data, Science

- Continuous Bag of Words (CBOW) takes context words as input to predict the target word.
- Skip-gram takes the target word as input to predict the context word

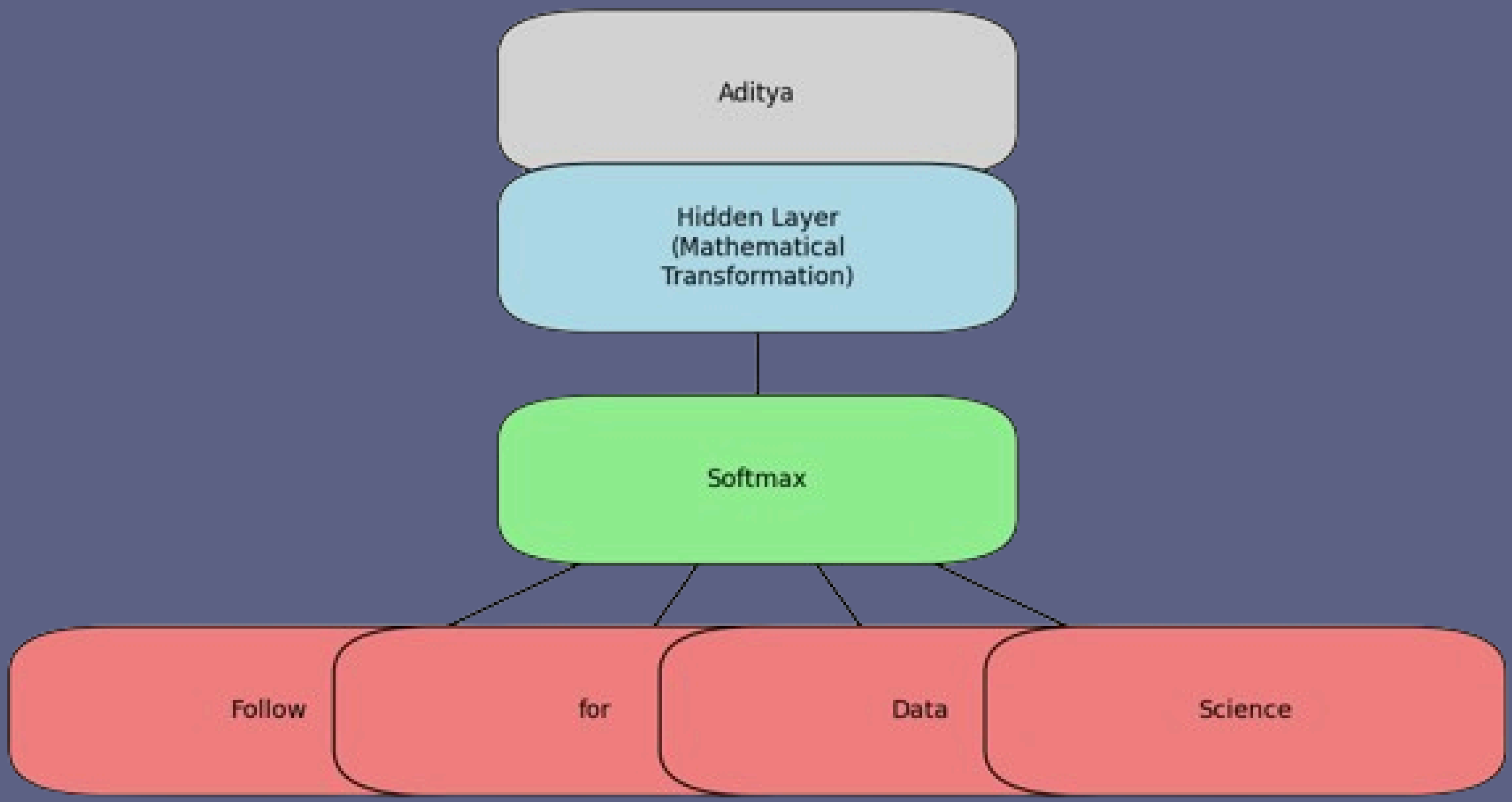
Continuous Bag of Words



Advantages:

- Faster training compared to Skip-gram.
- Works well with a large corpus.
- Good at predicting frequent words.

Skip-Gram



Advantages:

- Better at capturing complex word relationships.
- More accurate for infrequent words.
- Provides richer word representations.

Thank You!

> **Just as Word2Vec finds meaning in words through their relationships, we find purpose in life through our connections with others.**

Follow **Aditya Goyal** for more on DSA, ML, DL, NLP & Data Science