



Data Science Course



Lecture 27



Bag-of-Words

- Large sparse vectors are used to represent each word or to score each word within a vector to represent an entire vocabulary.
- Sparse because the vocabularies are vast and a given word or document is represented by a large vector comprised mostly of zero values.

Word Embeddings

- Words are represented by dense vectors.
- Words that have the same meaning have a similar representation.
- A vector represents the projection of the word into a continuous vector space.
- The position of a word within the vector space is learned from text.

Word Embeddings

- Two popular examples of methods of learning word embeddings from text include:
 - Word2Vec.
 - GloVe.