Text Representation

Problem statement
- → INPUT: Address of US presidents
- → TASK: Identify president from speech
  - → Often results in chunking of subproblems that are similar

BAG OF WORDS:
- → Orderless documentation representation, frequencies of words from a dictionary
- → Can often identify by taking "most used words"
- →

Histogram of Word Occurences

1-Hot Representation of a Word:

A [1 0 0 0 . . . . 0]ᵀ

Also [0 1 . . . . . 0]ᵀ

And [0 0 1 . . . . 0]ᵀ

Histogram is sum of one hot representation of all words

Weighted Words:
- → Not all are very useful
- → Some words like 'the, of, and, a' etc don't all value
  - ↳ known as stop words. (remove them from dictionary)
- → Some words are more important
  - → Words that occur multiple times
  - → Words that are unique to document

Term frequency (TF)

→ Higher freq, more relevant
→ Measures freq. of term in document
→ Specific to document
→ $TF(T) = \frac{\#\text{ of occurrences of } T \text{ in } D_i}{\#\text{ of words in } D_i}$

Inverse document freq. (IDF)

→ Rare/unique words
→ Need to weight down freq. terms while scale up rare ones

$$IDF(T) = \log_e\left(\frac{\#\text{ of documents}}{\#\text{ of docs with term } t}\right)$$

TF-IDF's

$IDF \times TF_{D1} = \begin{bmatrix} \\ \end{bmatrix}$ → $TF1DF1$

$IDF \times TF_{D2} = \begin{bmatrix} \\ \end{bmatrix}$ → $TF1DF_2$

Unigrams

↳ Stemming

Representing words

→ Words are atomic entities
→ One-hot representation doesn't capture meaning
  ↳ Words with similar/different meanings are equally apart
  ↳ Cosine distance is 0 b/c any two words
  ↳ Unlike apples & oranges

→ Can words be represented as vectors?
  → How do we capture meaning
  → How do we learn

Word2Vec:

→ Meaning of word captured by co-occuring words

→ Don't need meaning of word but rather collection of sentences

→ Rep. of middle word is sum of left & right words

Like a segment problem

$x_3 = f(w, x_1, x_2, x_4, x_5)$

↳ Learnable parameters / weights

$x_1 = R(a_1)$

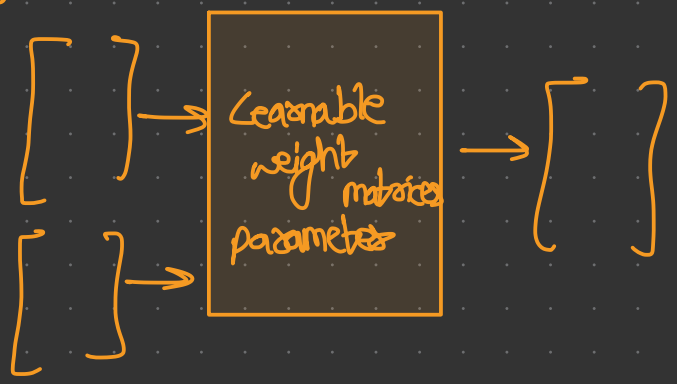
$x_2 = R(a_2)$

$x_3 = R(a_3)$

$x_4 = R(a_4)$

$x_5 = R(a_5)$

Summary:



Weight matrix: $V \times N$

Vocabulary size ↳ Size of representation

Dense; Semantic

$\begin{bmatrix} 2 \times 1 \end{bmatrix}$

Learned Matrix

$\begin{bmatrix} W^T_{(N \times V)} \end{bmatrix}$

↑ Precomputed

One hot; Sparse

$\begin{bmatrix} V \times 1 \end{bmatrix}$