Last session:

① Text Preprocessing

② Text Representation

- ① One Hot
- ② BOW
- ③ N-gram

* TFIDF
* word2vec

* Agenda:

① TFIDF

② word2vec

③ How to perform text classification using Machine Learning

→ Text pre

→ text representation

Tommorrow: RNN

① TFIDF → Term frequency - Inverse Document Frequency

Drawbacks → ① One Hot — sparce metrix

↳ Dimention issue

↳ sementic meaning not captured

↳ OOV

BOW → Sparce metrix

→ sementic meaning

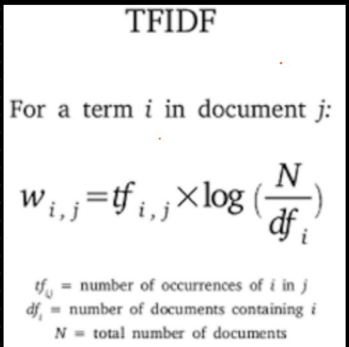
1 , 2 , 0

① This movie is Amazing! Amazing Perfonmance by SRK

— — — —

—

most frequent word will get mone weight →



TFIDF

For a term i in document j:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

0 ⇒

0.85 × 0.5

⇒ (- - ⇒

0.22 0.89

machine → 1.h

10 [illegible]

This	mai	is	Amz	perfo	by	SRK
0.85	0.78	0.5	0.99	0.7	0.4	0.3
—	—			—	—	—
	—		—		—	
					—	

→ postive

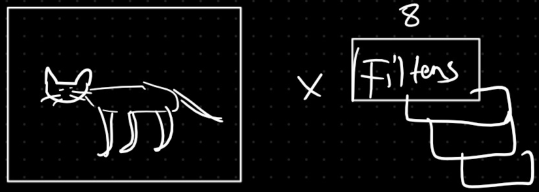
→ 100

Ann → weight →

② word2vec: It is using some kinds of NN to generate different features from a word

weights → ANN

CNN



□ → —

□ → |

□ → eyes

□ → Tail

I am the king in this Universe

NN →

BP

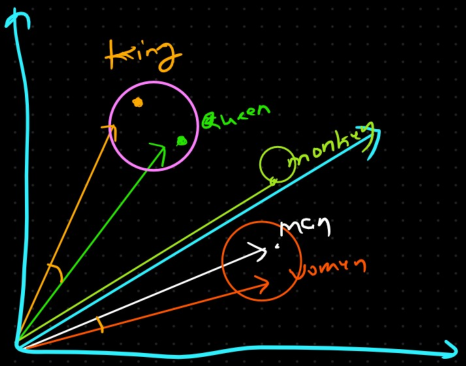
Featunes	King	Queen	Man	Women	Monkey
Junder	1	0	1	0	1
wealth	1	1	0.3	0.2	0
power	1	0.7	0.5	0.2	0
weight	0.8	0.5	0.7	0.5	0.3
speak	1	1	1	1	0

→ Word2vec

King ⇒ [1, 1, 1, 0.8, 1]

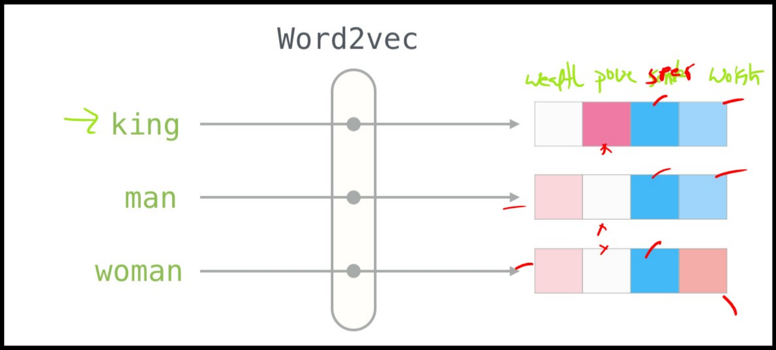
Queen ⇒ [0, 1, 0.7, 0.5, 1]

5D



substract

Featunes	King	Queen	Man	Women	Monkey
Junder	1	0	1	0	1
wealth	1	1	0.3	0.2	0
power	1	0.7	0.5	0.2	0
weight	0.8	0.5	0.7	0.5	0.3
speak	1	1	1	1	0



my name is Bappy, I am a Data

A Game Of Thrones Book One of A Song of Ice and Fire By George R. R. Martin PROLOGUE "We should start back," Gared urged as the woods began to grow dark around them. "The wildlings are dead."

Rev 8602

Simple preproces gensim

n 0 A Game Of Thrones Book One of A Song

s 2 of Ice and Fire By George R. R. Martin

l 3 PROLOGUE "We should start back," Gared

["A", "games"

gensim → word2Vec

custom → trained ← Data

Transfom

word2vect

100 →

[100,

king 100D

A Game Of Thrones Book One of A Song of Ice and Fire By George R. R. Martin PROLOGUE "We should start back," Gared urged as the woods began to grow dark around them. "The wildlings are dead."

→ 100

[~~Amazin~~, 100

~~This~~ -2

["This Book is Amazing, I read all the lessons. great Author",]

frequent Accuring count

con