Word Embedding

* Word Embeddings are type of word representation that allows words with similar meaning ~~for the~~ to have a similar representation.

* Word embeddings where individual words are represented as real valued vectors in a predefined vector space.

  * Each word → Mapped to → vector values.
    ↓
    tens of hundreds of dimensions.

  * Constrasted to thousands or million dims req for sparse word representation.

So each word is associated with a point in vector space. The no of features is less smaller than the size of the vocabulary.

→ Word embedding that is learned jointly with a neural network model on a specific NLP task.

→ Requires text to be cleaned and prepared such that each word is one hot encoded.

→ embedded layer is used on the front end of the neural network and is fit in a supervised way.

(1) Word2vec is a statistical method for efficiently learning a standalone word embedding for text corpus

by Tomas Mikolov (Google) 2013

* This representation surprisingly good at capturing syntactic and semantic regularities in language.

King - man + woman = Queen

→ CBOW Continous Bag of word model

→ Continous Skip Gram model

CBOW model learns the embedding by predicting the current word based on context.

Continous SkipGram model learns by predicting the surrounded words given a current word

* context is defined by a window of neighbouring words
* window is a parameter

(2) ~~Glove~~

② Glove : Global vector for word representation is an extention of word2vec for efficiently learning word vectors by Pennington + Stanford

TASK

Date_____

* Word embedding :- *

It is a technique which converts word into vectors

- Word Embedding
  - Count or Frequency
    1. BOW
    2. TF-IDF
    3. one hot encoding
  - Deep learning Trained models
    1. Word2Vec
       - CBOW
       - Skip Grams.

Embedding layer

Text → Vec

Word2Vec :- Feature Representation

f_1	f_2	f_3	f_4	f_5	... f_n
Boy	Girl	king	Queen	Apple	Mango

↓ Vocabulary

* every word will be converted to vector
* limited dimensions
* sparsity will be reduced (not find many zeros)
* Semantic meaning will be maintained

drawbacks of BOW, TF-IDF

Date _____

(1) { honest good } *** Symantic relationship

(2) Sparse Matrix
→ huge dimensions

	f_1	f_2	f_3	f_4	f_5	f_6
	Boy	Girl	king	Queen	Apple	Mango
Gender	-1	1	-0.92	+.93	0	0
Royal	0.01		0.95	0.96		
Age	0.03					
Food						
↓						
300 Dimen						

→ Trained model
→ Semantic relationship
→ related vector

TASK

Date _____

King [0.96 0.95]

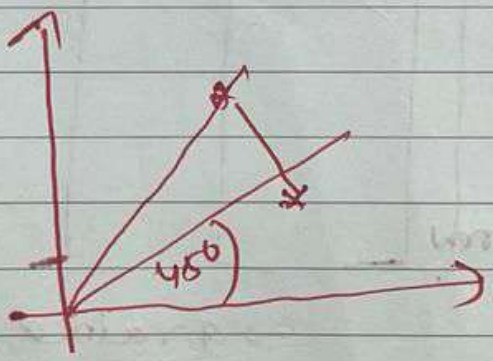
Queen [-.96, 0.95]

Man [0.95 0.98]

Women [-0.94, -0.96]

King - Man + Woman = Queen

Cosine Similarity



Distance = 1 - Cosine Similarity

$\cos 45^\circ = 0.7071 = \frac{1}{\sqrt{2}}$

$\Rightarrow 1 - 0.7071$

$\Rightarrow \simeq 0.29$

more towards 1 very similar

* Word2vec

① CBOW → {Continous Bag of words}

Corpus : dataset Training

[krish channel is related to data science] → total 7 words

Window size = 5

* window must be odd no
Bigger window → good model

Independent feature O/P

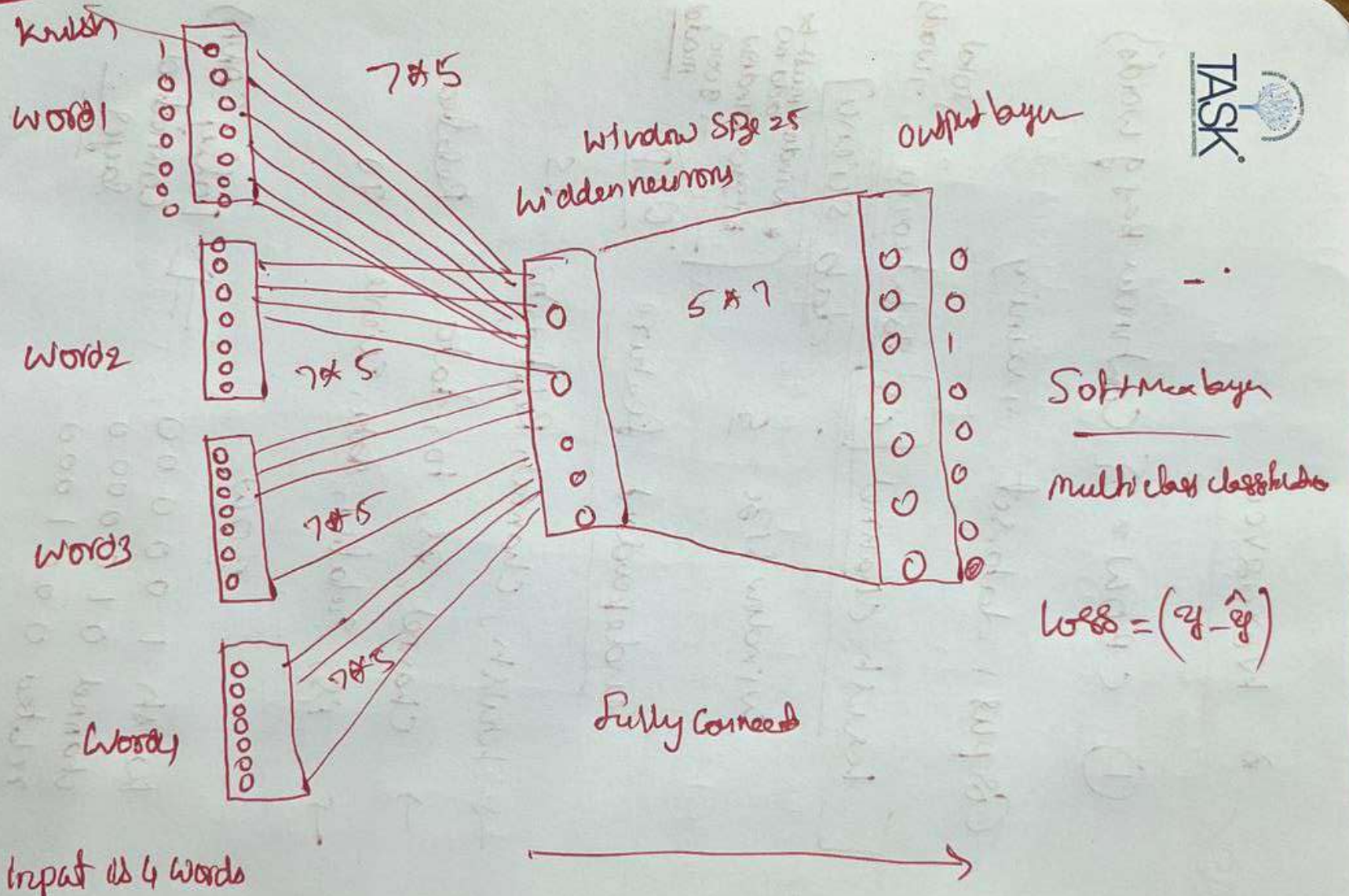
→ krish, channel, Related, to Is
→ channel, Is, to, data Related
→ Is, related, data, science To

BOW

krish 1 0 0 0 0 0
channel 0 1 0 0 0 0
related 0 0 0 1 0 0

⇒ Fully (ANN) Connected layer

neurons



SoftMax layer

multi class classification

$loss = (y - \hat{y})$

TASK

Date_____

② Skip grams

I/P	O/P
is	krish, channel, related, to
Related	channel, is, to, Data
to	is, related, Data, Science