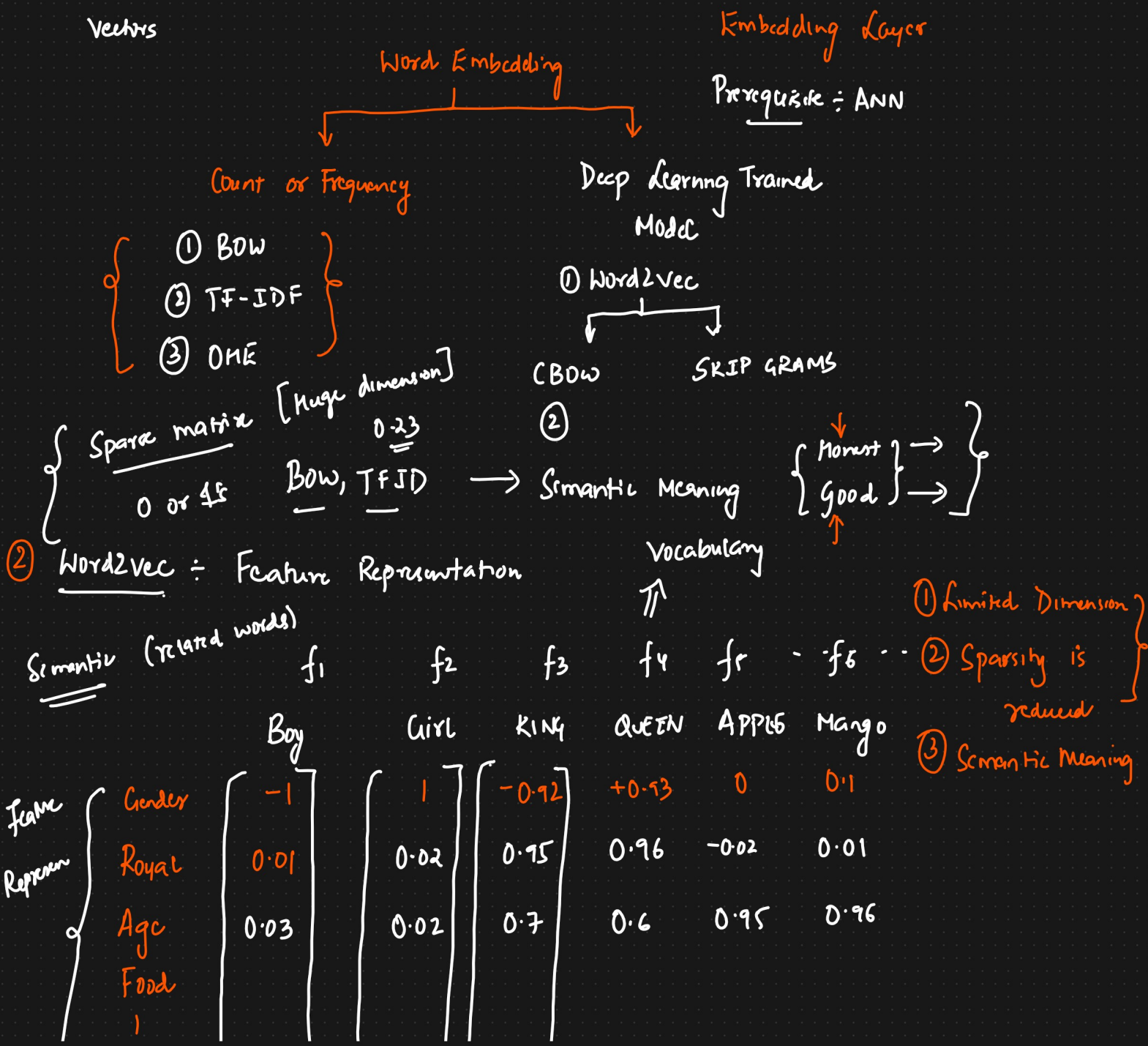


Day 4 → NLP

- ① Word Embeddings
- ② Word2Vec → CBOW (Continuous Bag of Words)
→ Skipgram
- ③ Practical Implementation Using Python

① Word Embeddings → It is a technique which converts words into Vectors



300 dimension [] [] []

$$\text{KING} - \text{Man} + \text{Women} = \underline{\underline{\text{Queen}}}$$

$$\text{King} [0.96 \ 0.95]$$

$$\text{Man} [0.95 \ 0.98]$$

$$\text{Queen} [-0.96, 0.95]$$

$$\text{Women} [-0.94 \ -0.96]$$



Matching

Cosine Similarity

$$\text{Distance} = 1 - \text{Cosine Similarity}$$

$$= 1 - 0.7071 = \underline{\underline{\approx 0.29}}$$

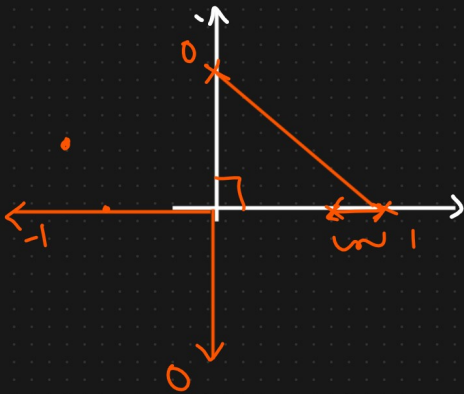
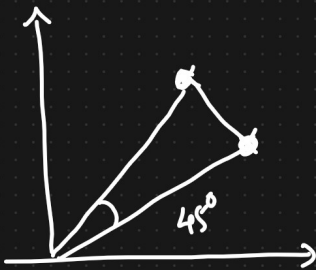
Eulerion Distance

$$\text{Cosine-Sim} = \cos \theta$$

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

$$\cos - \text{Sim} = \cos 90 = 0$$

$$\text{Dist} = 1 - 0 = \underline{\underline{1}}$$



⊛ Word2vec

(i) CBow { Continuous Bag of Words }

CORPUS ÷ DATASET TRAINING

[KRISH CHANNEL IS RELATED TO DATA SCIENCE]

Context Window Size = 5

TRAINING DATA

Word2vec

Text → Vectors of Semantics

Independent feature I/O/P

O/P I/P

BoW

→ KRISH, CHANNEL, Related, To

IS

KRISH

1 0 0 0 0 0

→ CHANNEL, IS, To, DATA

Related

Channel

0 1 0 0 0 0

To

IS

0 0 1 0 0 0

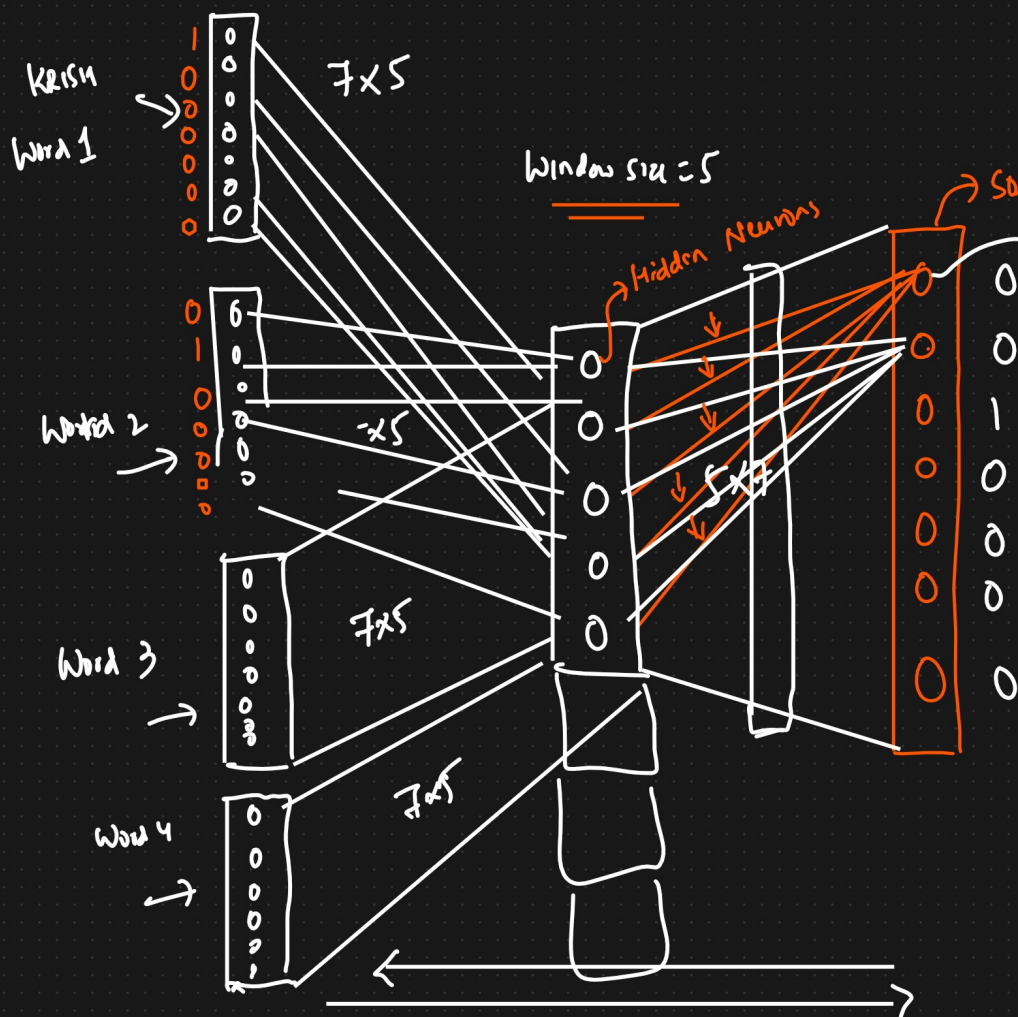
→ IS, Related, DATA, SCIENCE

Related

0 0 0 1 0 0

C BoW

(ANN) Fully Connected Layer



$$\text{Loss} = (y - \hat{y})$$

→ KRISH

Window size

[0.92 0.94 0.24 0.36 0.45]

Window size = 5

Window size = 300

300

② Skip GRAM

I/P

O/P

IS

KRISH, CHANNEL, Related, To

Related

Channel, IS, To, Data.

To

IS, Related, Data, Science