

Cosine Similarity

Purpose :- Calculate the distance b/w vectors

Cosine Similarity { High }
Cosine Distance { Low }
Inverse relationship

Formula :-

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \times \|B\|}$$

{ A, B }
vectors

(\cdot) = Sum of element wise multiplication

$\|A\|$ = length / Magnitude of the vector $A = \sqrt{\sum a_i^2}$

$$\|B\| = \sqrt{\sum b_i^2}$$

Person A

Person B

	Action Rating	Romance Rating	Vector
Person A	5	3	$[5, 3]$
Person B	10	6	$[10, 6]$
Person C	1	5	$[1, 5]$

Step 1 :- Dot Product

$$A \cdot B = (5 \times 10) + (3 \times 6) = 68$$

Step 2 :- Calc the Magnitudes

$$|A| = \sqrt{5^2 + 3^2} = \sqrt{34} \approx 5.83$$

$$|B| = \sqrt{10^2 + 6^2} = \sqrt{136} \approx 11.6$$

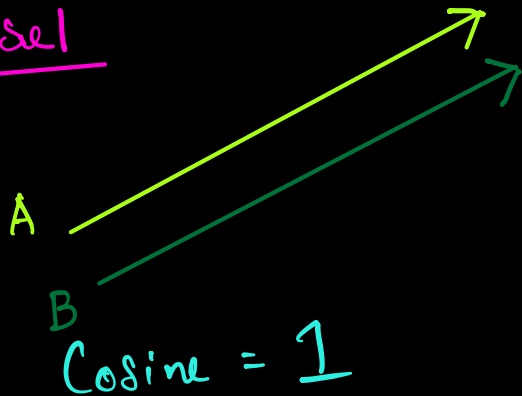
Step 3 :- Apply the formula

$$\begin{aligned} \text{Cosine Similarity} &= \frac{68}{5.83 \times 11.6} \\ &= \frac{68}{68.03} \approx 1 \end{aligned}$$

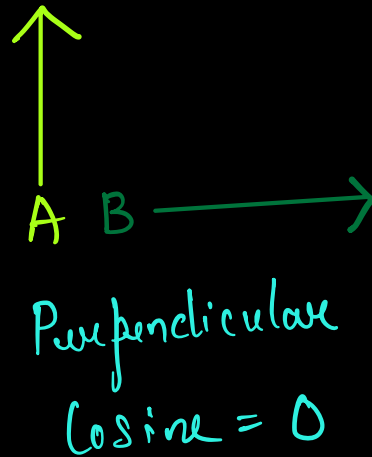
A and B are nearly identical.

Vector Visual

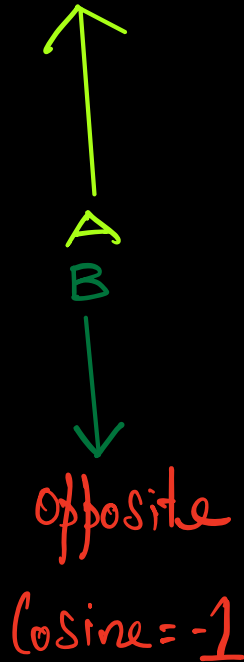
Case 1



Case 2



Case 3



Value Range = $[-1, 1]$

Person A to Person C

Step 1 :- Dot Product

$$A \cdot C = (5 \times 1) + (3 \times 5) = 20$$

Step 2 :- Magnitude

$$||C|| = \sqrt{1^2 + 5^2} = \sqrt{26} \approx 5.1$$

Step 3 :- Apply the formula

$$= \frac{20}{5.83 \times 5.1} = \frac{20}{29.72} = 0.67$$

Person A & Person C have different taste

1) Euclidian Distance (L2)

Formula:- $\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots}$

1) Straight Line Distance

* Manhattan Distance (L1)

✓ Sparse vectors

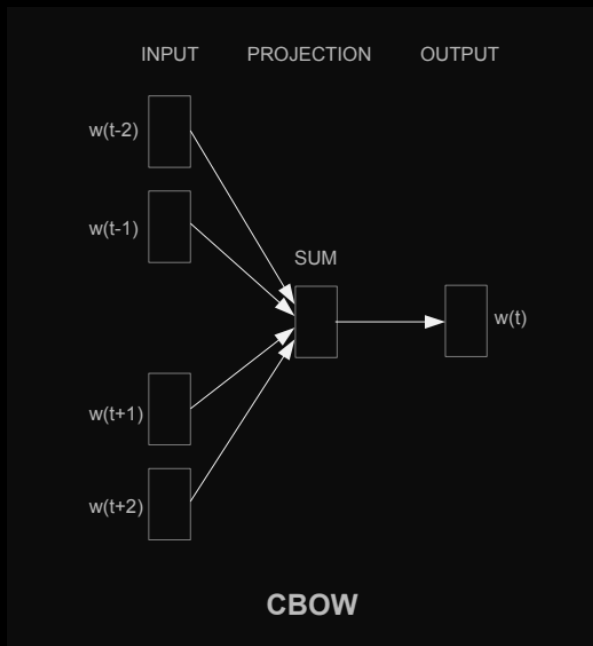
Word 2 Vec

word embedding technique

Two Architectures

1) CBOW

Continuous Bag of Words



2) skip-gram

Cranberry

is

best

fruit

the

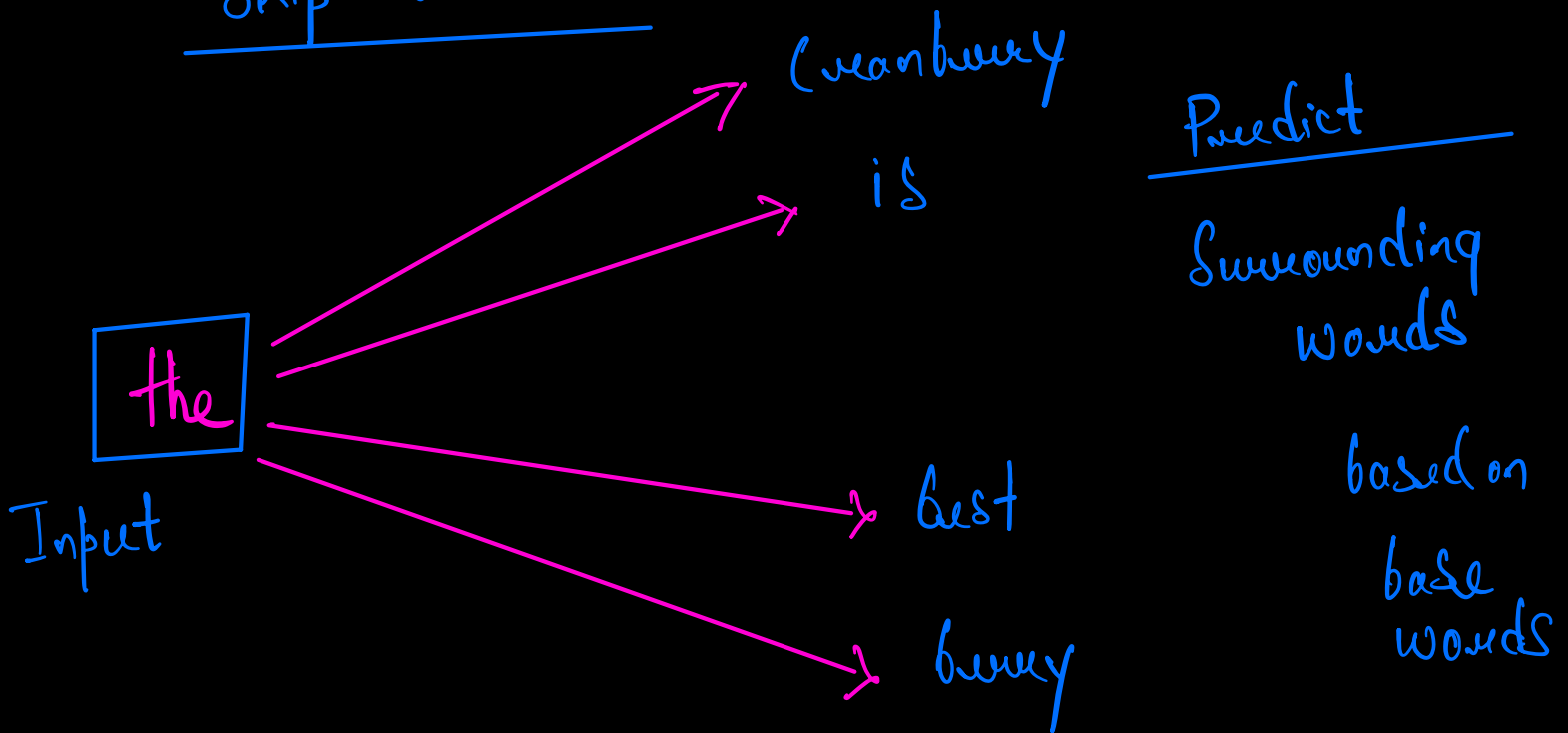
→ Target

* Surrounding context words

Cranberry is the best fruit.

- 1) Very fast to train
 - 2) Good with frequent words.
- X for rare word X

Skip-gram



Benefits of Skip-gram

- 1) Rare words

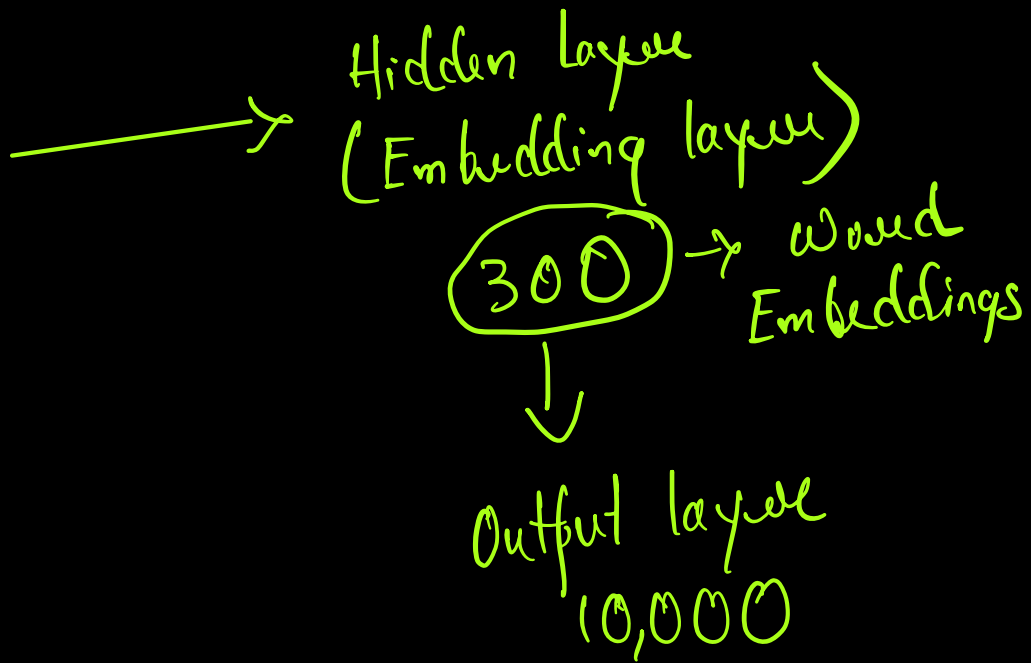
Con:

Slower to train as compared to CBOW

Week 2 Lec (2013)

Shallow Neural Network

Input layer
(Vocabulary Size)
10,000



Context Window

My name is Paul and I live in Chennai.

```
graph LR; S[My name is Paul and I live in Chennai.]; subgraph Size2 [Size: 2]; S --> W1[is Paul and]; S --> W2[Paul and I]; end; subgraph Size3 [Size: 3]; S --> W3[is Paul and I]; S --> W4[Paul and I live]; end;
```

Size: 2

Size: 3

Training Set Sample

(CBOW) (target word, context words) window=2
the {cucumber, is, best, honey}

(Skipgram)