

You are screen sharing EN Stop share

z Webinar Chat

Ashish Nautiyal to Hosts and panelists
AN no

Bharath Muthusivam to Hosts and panelists
BM not solved

Urmil Shah to Hosts and panelists
US cant tell

Ashish Nautiyal to Hosts and panelists
Who can see your messages? Recording on

To: Hosts and panelists

Type message here...

R1 :- This pasta is tasty → positive ↪

R2 :- This pasta is hot tasty → Negative ↪

Stop words

$$\begin{array}{c} R1 [1 \quad 1] \\ R2 [1 \quad 1] \end{array} v_1 = v_2$$

R1 :- Pasta tasty → both are same
 R2 :- Pasta tasty → R1 = R2 → OOV

→ Sparsity

→ OOV

* Don't apply stop words

Vocab :-

	This	pasta	is	tasty	not	
R1	1	1	1	1	0	[11110]
R2	1	1	1	1	1	[11111]

Which word is more significant?

n-grams - uni-gram (Bow)

n-gram range = (1,1) — Bag of words

$n = 1 = \text{Bow}$

$n = 2 = \text{bi-grams} = \text{pair of words}$

$n = 3 = \text{tri-grams} \rightarrow 3 \text{ consecutive words}$

Data augmentation

10 words
 \rightarrow 100 words

n-grams

R1: This pasta is tasty

R2: This pasta is not tasty

vocab :- bi-grams = pair of words

	This	pasta	is	is	tasty	
R1	1	1	1	0	0	$[1110] v_1$
R2	1	1	0	1	1	$[1101] v_2$

$R_1 \neq R_2$

$V_1 \neq V_2$

R1 This pasta is tasty.
R2 This pasta is not tasty.

Vocab

This pasta is not this pasta pasta is tasty

This pasta is pasta is tasty pasta is not if

This pasta is tasty This pasta is not pasta is
14 15 16

zm Webinar Chat

Virat to Hosts and panelists
V can you plz repeat it again
pair of words
clear sir

Bharath Muthusivam to Hosts and panelists
BM how to decide which n-gram to use? like bi or tri or tetra?

Ashish Kumar to Hosts and panelists
AK yes sir

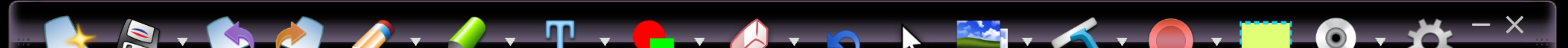
Ashish Nautiyal to Hosts and panelists
AN Yes

Bharath Muthusivam to Hosts and panelists
BM yes sir

Who can see your messages? Recording on

To: Hosts and panelists

Type message here...



You are screen sharing

EN Stop share

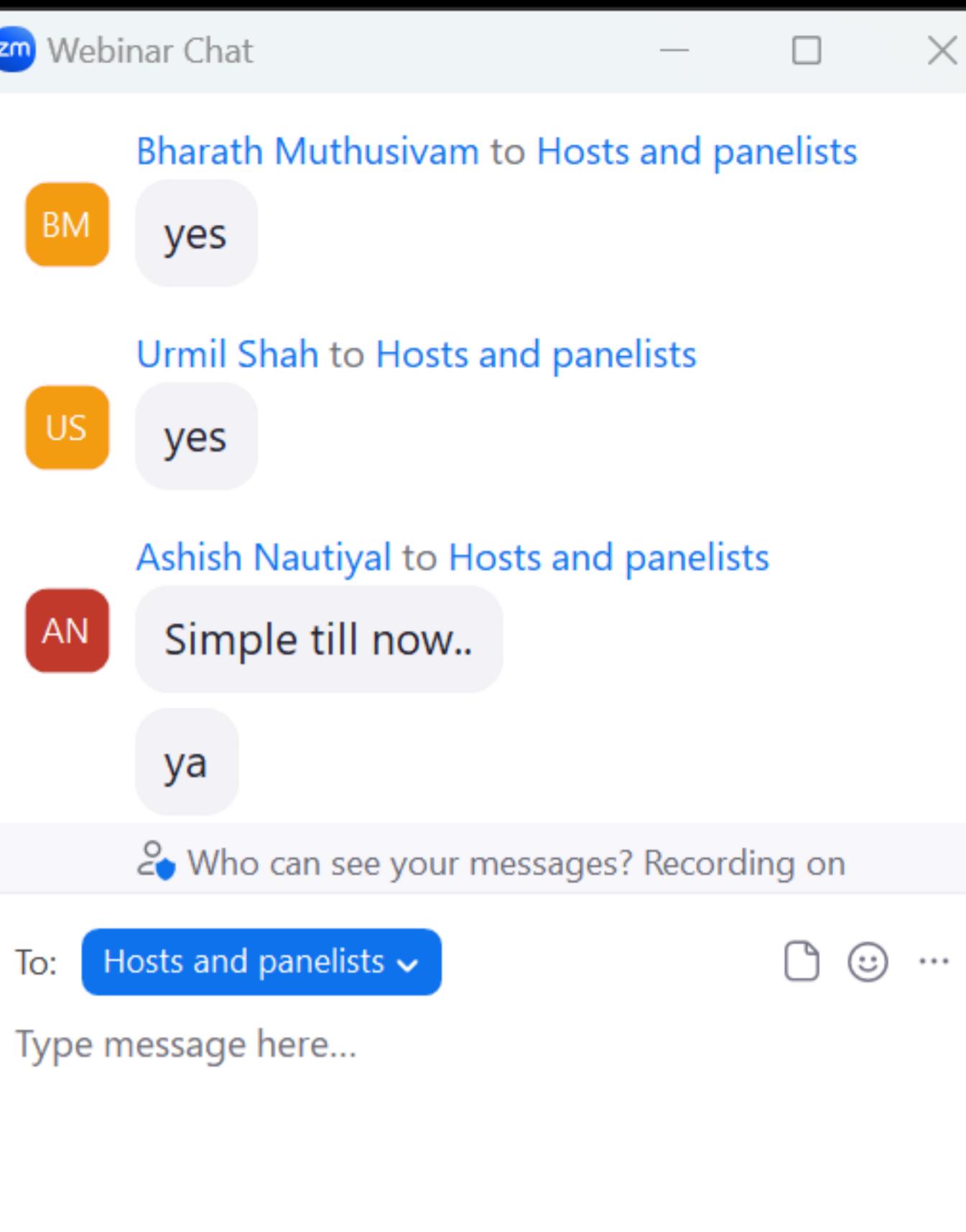
Talking: Learnvista Private Lim...

TF-IDF - Term Frequency - Inverse Frequency

$$TF = \frac{\text{No. of rep of words in sentence}}{\text{No. of words in sentence}}$$

TF = Probability

$$0 \leq TF \leq 1$$

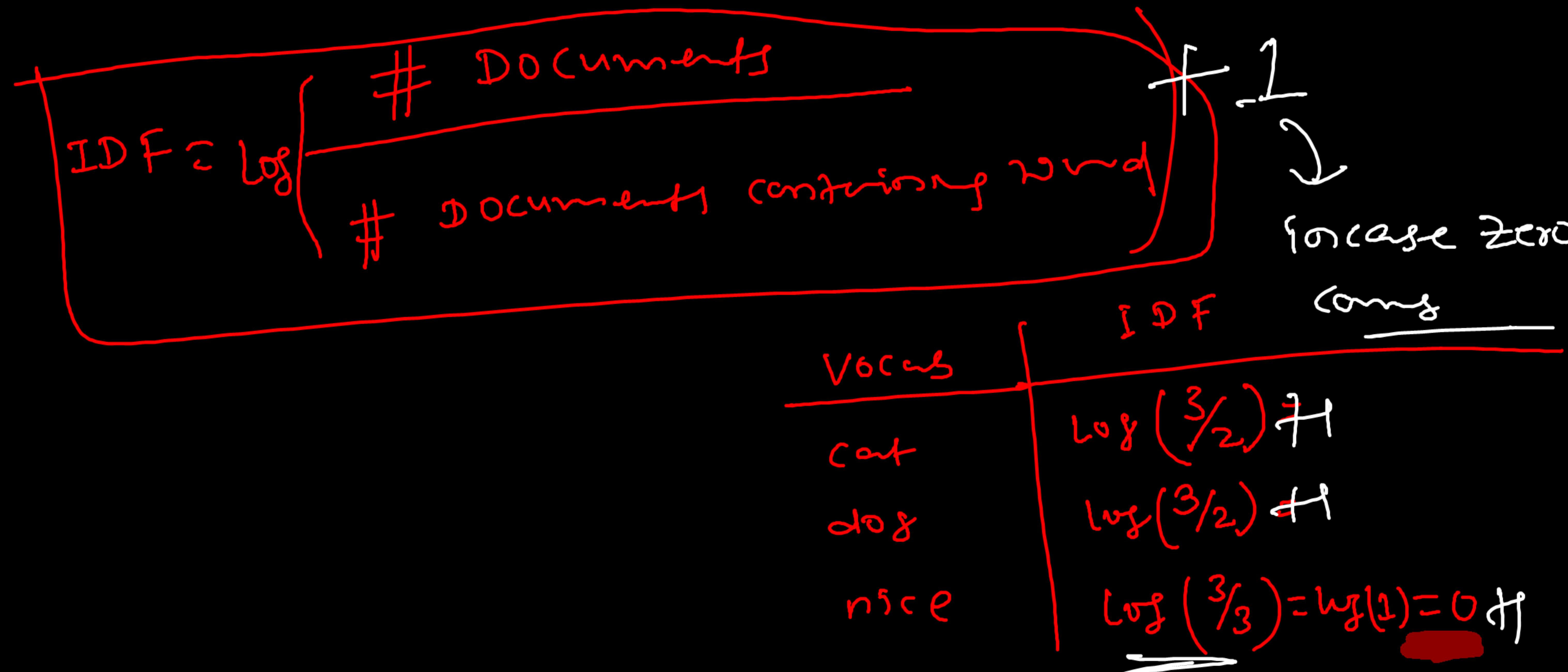


Sent 1 :- nice cat
nice cat
Sent 2 :- nice dog
nice dog
Sent 3 :- cat dog nice
cat dog nice

Vocabs	S1	S2	S3
cat	1/2	0	1/3
dog	0	1/2	1/3
nice	1/2	1/2	1/3

$$TF = \frac{\# \text{ of rep of words in a sent}}{\# \text{ of words in sent}}$$

IDF → Inverse Documents frequency



TF * IDF

	F1	F2	F3
cat	0 + 1	0 + 1	0 + 1
dog	0 + 1	0 + 1	0 + 1
nice	$\frac{1}{2} \times 0 = 0$	$\frac{1}{2} \times 0 = 0$	$\frac{1}{2} \times 0 = 0$

free

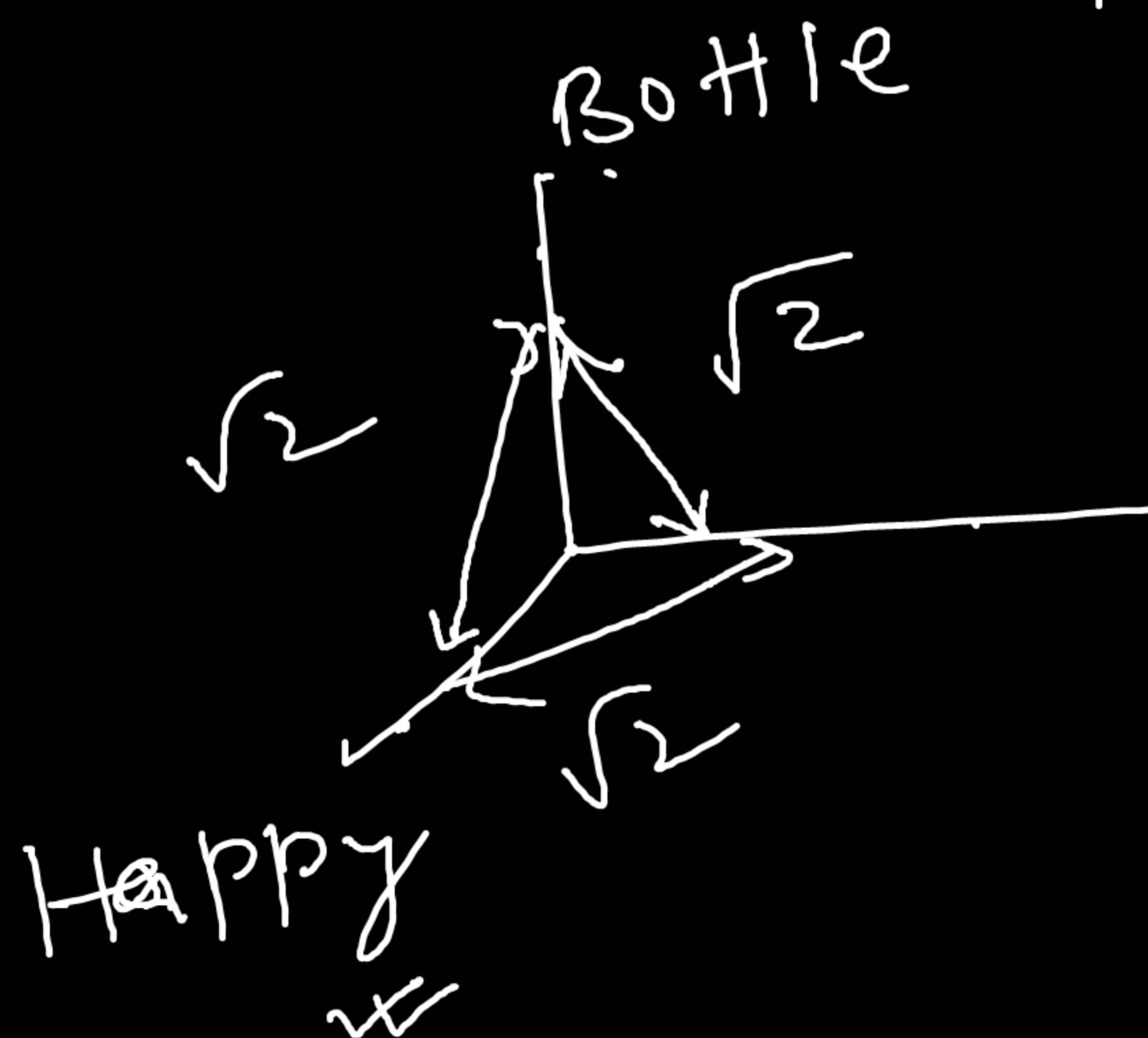
$$\text{cat} = 2$$

$$\text{Dog} = 2$$

$$\text{nice} = 3$$

Adv

- Better informs
- n-gram



Dis-adv

→ Sparsity ✓

→ OOV

→ Semantic meaning

Dense

OOV

glove

