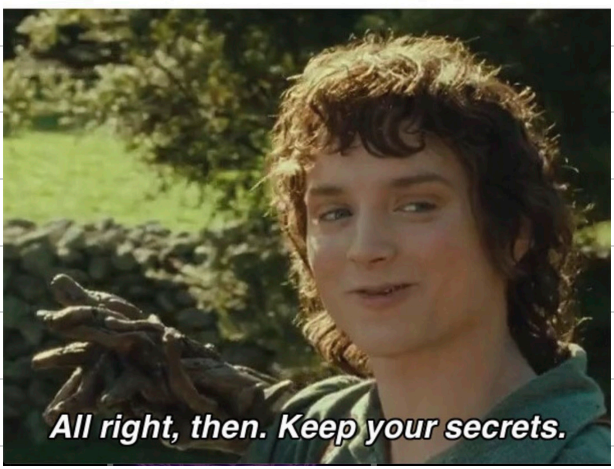
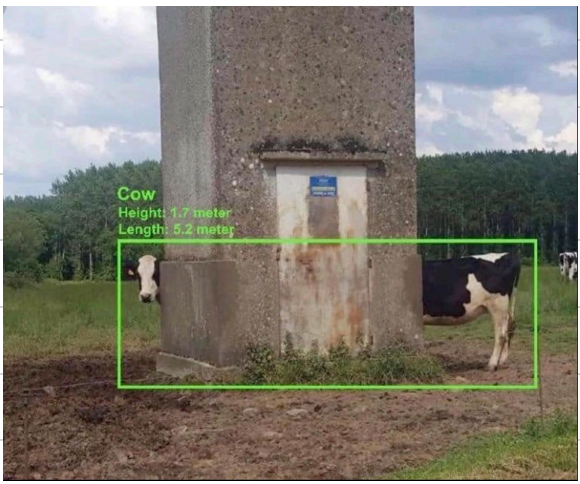
When a paper doesn't describe their approach or link their code and only has benchmarks





Agenda

1. POS Tagging - What and How?
2. Topic Modelling - LDA

Nobody:
Me to Text to Speech model for no reason

LATENT DIRICHLET ALLOCATION

2003

50% Topic 1
50% – Topic 2

Doc1 w_{11} w_{12} w_{13} — —

Doc2 w_{21} w_{22} w_{23} — — → 100% → T_1

Doc3 w_{31} w_{32} w_{33} — —

→ 10% T_1, 80% T_2

→ Problems with LDA → BOW model.
- ↳ no context
- ↳ no Sequence
- ↳ Good/Great → complete seperate

→ Value to choose no of topics 12
Eg: 3

100% → topics ear

Doc 1 →

→ LDA works at a word level

$D_1 \rightarrow$ 10 words

6 $\rightarrow T_1$
4 $\rightarrow T_2$

60% $\rightarrow T_1$
40 $\rightarrow T_2$

$\simeq$ Soft K-means Clustering

Sports — ["Football" x0.4, "Fifa" - 0.2]

→ Two Ways to measure LDA

① Doc / Topic matrix → Doc1 → 60% T_1
40% T_2

② Topic / Word matrix

Bank → Finance
→ Environment

Apple → Tech
→ Food

Algorithm

① Allot any topic randomly to all the words

2

→ N iterations

For Each document:

For each word in doc w:

For each toPic

1. Ratio of word w → t_i / All the words in t_i | $P(w/t_i)$
2. Proportion of words in t_i for a document | $P(t/d)$
3. Multiply

$$P(w/t_i) \times P(t/d)$$

Example:

toPic 0 toPic 1 | K = 2

- **Doc1**: orange - 50, soup - 50, diet - 20, concert - 30
- **Doc2**: apple - 40, health - 10, guitar - 50, music - 60
- **Doc3**: banana - 45, bread - 55, song - 25, festival - 35

Document / Topic matrix

	Topic 0	Topic 1
Doc1	70	80
Doc2	110	90
Doc3	80	80

Topic word matrix

	Topic 0	Topic 1
orange	50	0
soup	0	50
diet	20	0
concert	0	30
apple	0	40
health	0	10
guitar	50	0
music	60	0
banana	45	0
bread	0	55
song	0	25
festival	35	0

260 210

50

50/260 70/150

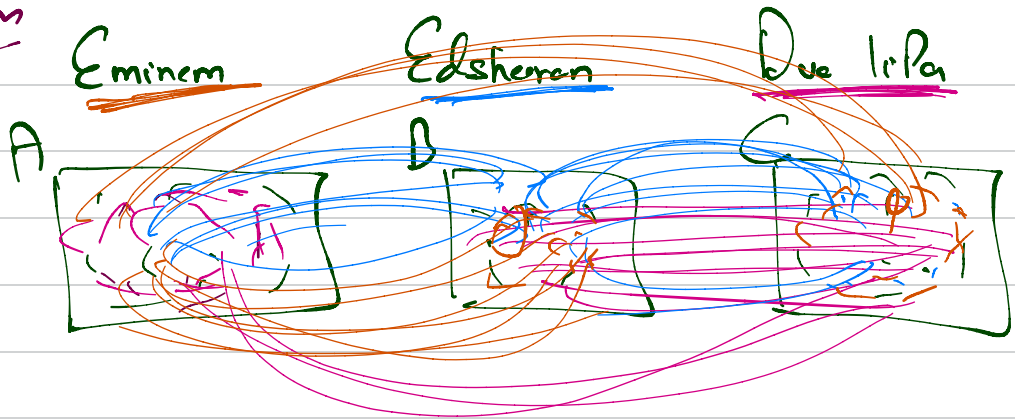
① $P(\text{orange}/t_0) \times P(t_0/d)$ = 0.089

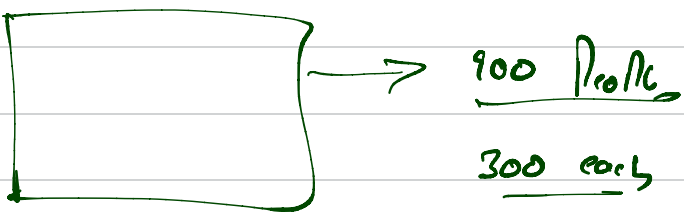
② $P(\text{orange}/t_1) \times P(t_1/d)$ = 0.10.

$\frac{50}{260} \times \frac{50}{150}$

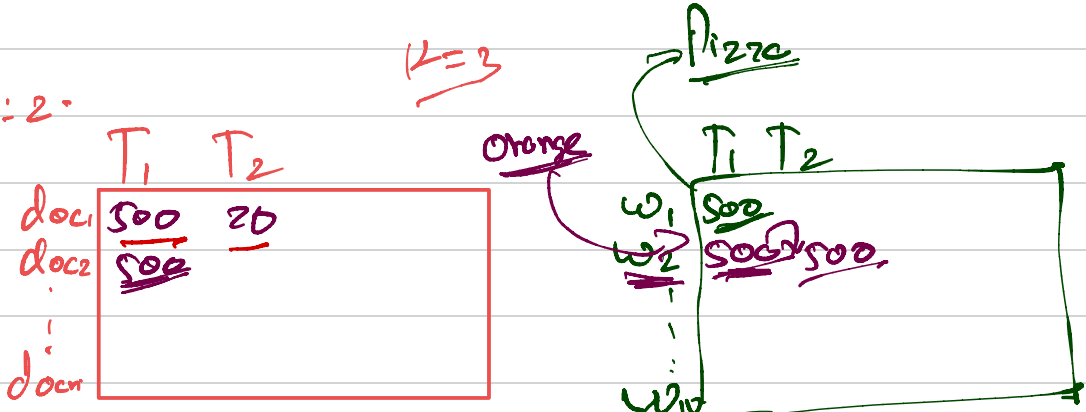
01 → Topic 1

8:30 pm





Eg:2.



20

500/580

$P(\text{cheese} \mid t_1) \times P(t_1 \mid D_1)$

20/580

$P(\text{Cheese} \mid t_2) \times P(t_2 \mid D_1)$

$= 1/2$

$P(\text{Orange}/t_1) \times P(t_1/D_1)$ — Highlighted

$P(\text{Orang}/t_2) \times P(t_2/D_1)$

1

Allotted to T_2

Rel small

Ex: 3

		500	200	5	3
D_{OC1}:	Mac	APPle	orange	mouse
D_{OC2}:	CV	ball	ML	

T_1

T_2

500

2

5

	T_1	T_2
D_1	505	203
D_2	505	2

	T_1	T_2
Mac	500	0
APPle	0	200
Orange	5	0
mouse	0	3
CV	500	0
ball	0	2
ML	5	0

Mac:

T_1: $P(mac(t_1) \times P(t_1/d_1)$ — $P(mac(t_1)$: 500/1010; $P(t_1/d_1)$: 505/708

$= 0.352$

T_2: $P(mac/t_2) \times P(t_2/d_1)$ — $P(mac/t_2)$: 505/705; $P(t_2/d_1)$: 703/705

$= 0.71$

Mac → T_2

	T_1	T_2
D_1	5	703
D_2	505	2

	T_1	T_2
Mac	0	500
APPle	0	200
Orange	5	0
mouse	0	3
CV	500	0
ball	0	2
ML	5	0

APPle → $P(APP/t_1) \times P(t_1/d_1) = 0.11$ (200/710 × 205/507)

$P(APP/t_2) \times P(t_2/d_1)$

200/706 × 203/208 = 0.28