07.09.24
Thursday

Statistical NLP (Text → numbers)

1) TF - IDF (Term freq & Inverse Document-freq)

$$TF = \frac{\text{No. of terms in the document}}{\text{Total no. of terms in the document}} = \left(\frac{x}{N}\right)$$

$$IDF = \log\left(\frac{\text{number of documents}}{\text{no. of documents where term has appeared}}\right) = \left(\frac{D}{T}\right)$$

∴ now [TF-IDF = TF * IDF]

D_1 : I'm reading a newspaper. I'm also reading a book.

D_2 : I'm reading a magazine. It has events

D=2, term = reading, N =

TFIDF (reading) ?

TF(reading, D_1) =

$TF(reading, D_1) = \frac{2}{10}$

$IDF(reading) = \log\left(\frac{2}{2}\right) = 0$

$TF(newspaper, D_1) = \frac{1}{10}$

$IDF(newspaper) = \log\left(\frac{2}{1}\right) = TF * IDF = \left(\frac{1}{10}\right) 0.3$

note:
if a term present in all doc then its not relevant or important in one content.

» Feature Extraction

Tw1 : I'm happy because I'm learning NLP
Tw2 : I'm happy

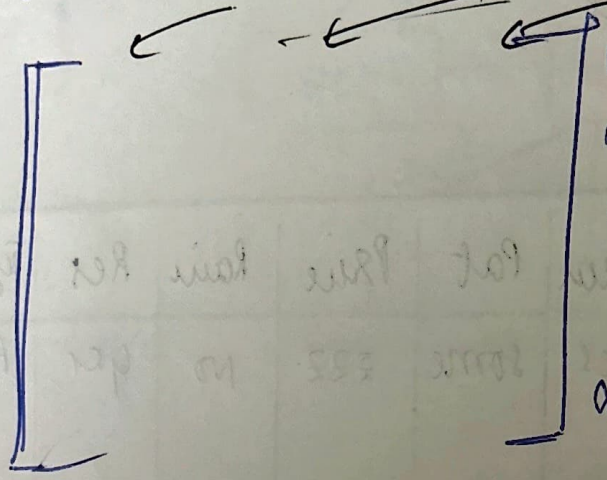
Tw2 : I'm sad, I'm not learning NLP
Tw2 : I'm sad

Row 3
Row 3
⋮

Row m
Row m

→ x_3
⋮
→ x_m.

* First column is Bias, Positive, Negative.



$(x^m \times 3)$ is the dimension of the matrix

1) Bag of words → Bow.
2) word to vector → skipgram.
3) Glove
4) BERT
5) LLM based

* Unit V *

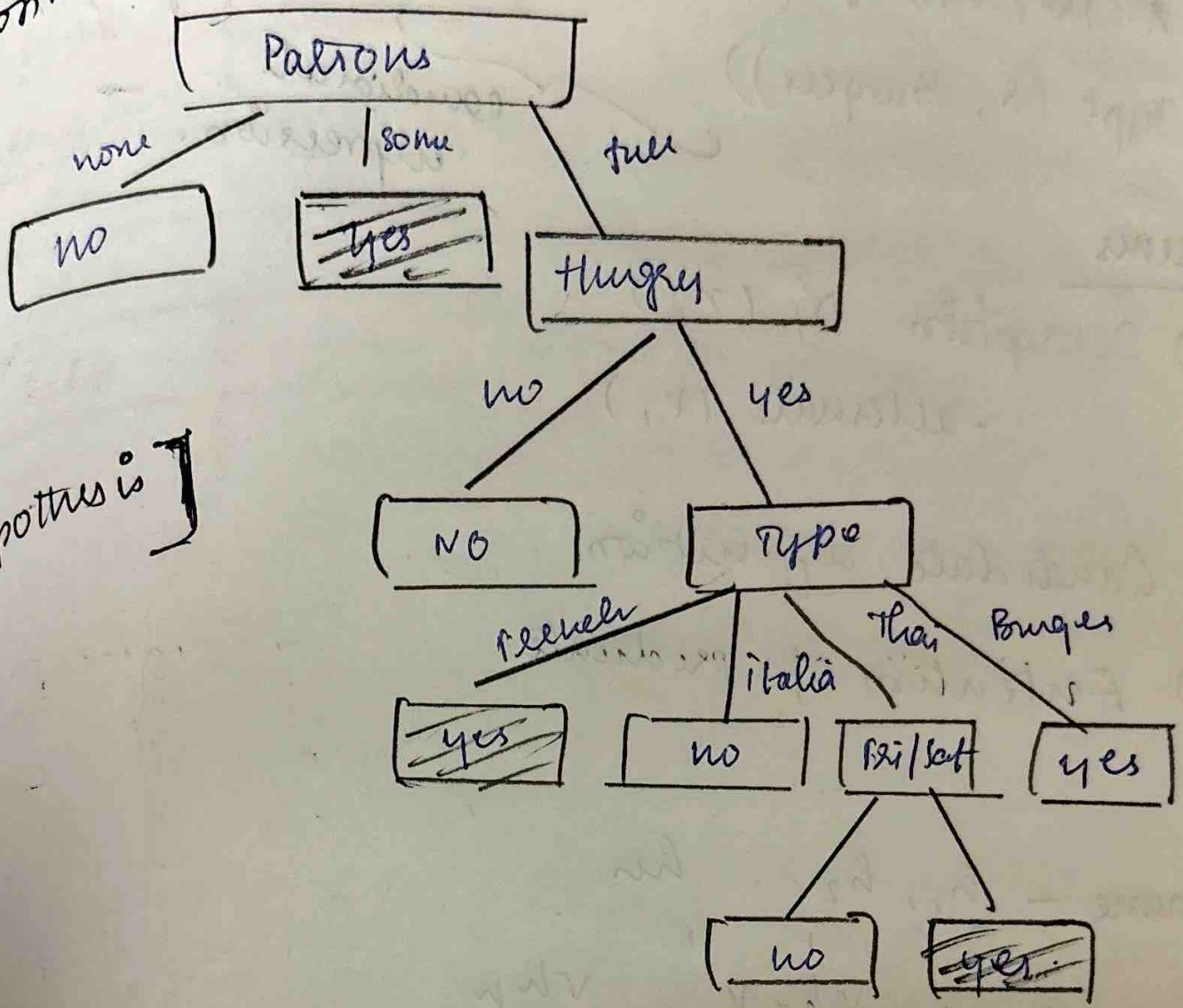
1) Logical formulation in learning

2) Application of NLP - chatbot

Chapter-18 Decision Tree

Ex	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est-
x_1	yes	no	no	yes	some	$$$	no	yes	French	0-10
x_2										
.										
.										
.										
.										
x_{12}										

decision tree



Hypothesis as logical sentence

i) alternate(x,) ∧ ¬bar(x,) ∧ ¬Fri/Sat(x,) ....

2) classification - willwait(x,) ∨ ¬willwait(x,)

Hypothesis

$\forall x \; Goal(x) \iff C_j(x)$ → candidate expression.

$\forall x \; willwait(x) \iff Patrons(x, some) \vee$
$(Patrons(x, full) \wedge Hungry(x)$
$\wedge Type(x, French)) \vee (Patrons(x, full)$
$\wedge Hungry(x) \wedge Type(x, Thai)$

$\wedge$ (Res/Sat(x) $\vee$ (Patrons(x, f(x)) $\wedge$ Hungry(y)) $\wedge$ Type(x, Burgers))

candidate expression.

Terms

(1) Description $D_i(x_i)$

- alternate (A_i)

(2) Candidate definition

(3) Extention of predicate.

Assume – $h_1, h_2 \ldots h_n$

$h_1 \vee h_2 \vee h_3 \vee \ldots \vee h_n$

↳ logically consistent

↳ logically inconsistent

↳ FN

↳ FP

Actual	Predicted	
True	True	→ True positive
False	False	→ True negative
True	False	→ False negative
False	True.	→ False positive.