

Coding stuff

NLP:- word embeddings:-

Sept 13 NER Tagging — Supervised Learning
classification - discrete.
kernel - SVM
KNN - 0 training time
Simple clusters — parametric unsupervised learning
image caption - multi model
Active learning
Cross validation.
Ensemble learning
multiple models

→ NLP finishing + NLTK Book.

→ chapter 6 ←
classifying Text. } pipeline.

* Gender Identification : } unethical?

Ex:- 10,000 docs of Indian parliament
classify into financial / not. } get a feature vector

Bag of Words model.

→ frequency & unique words in feature vector.

→ Ex: classification → Democratic Bow's
→ Republic Bow's

→ Exploiting Context } → ngrams?

Sequence classification: NER, Coreference resolution.

define a dozen?

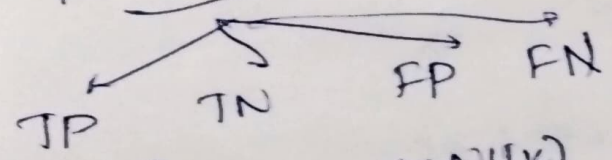
Ex: HMM (HMM model)
Markov model } prob of one depends on the before word...

RNN → difficult for long sentences

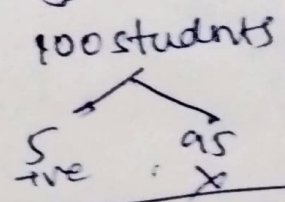
LSTM ← find entities & then clustering

→ coreference resolution: (corenlp.run) ← Stanford

→ Evaluation on Test model.eval()
precision vs recall ← (~~accuracy~~)



(CS635 slides in ch6 NLPK)



$$P = \frac{TP}{TP+FP} \quad R = \frac{TP}{TP+FN}$$
$$F_{score} = \frac{2PR}{P+R}$$

→ Confusion Matrix:

Binary ← confusion matrix is for multi classification.

Ch-8: Ubiquitous Ambiguity.

CFG: Context-free grammar
(Shift reduce parser)
Inconvenient assumption } no subtree in a grammar tree is connected with another subtree

Discrete: Idea of a Topic model.

doesn't care about the order.
500 queries } → 10 list of topics [...]

