Overfitting

Acc

100%

Training Set Acc

Test Set Acc.

Regularization →

← Model Complexity

In-text tagging

- Previous: Text Classification

  - Input: $\vec{x} = (\text{"I"}, \text{"love"}, \text{"this"}, \text{"movie"})$

  - Output: $y = \text{POS\_SENT}$

- Next: Token Tagging (Classif.)

  - Input: $\vec{x} = \begin{pmatrix} I & love & this & movie \end{pmatrix}$

  - Output: $\vec{y} = \begin{pmatrix} PRO & VERB & DET & NOUN \end{pmatrix}$

- Let's move to classifying **within the text!**

  - Tasks you can do yourself, with the right heuristics or logistic regression features (or other NLP models)

  - Do it with a pretrained, off-the-shelf system as part of a larger system, especially for syntactic/semantic linguistic analyses

3

Part of speech tagging

- I saw the <u>fire</u> today
  NOUN

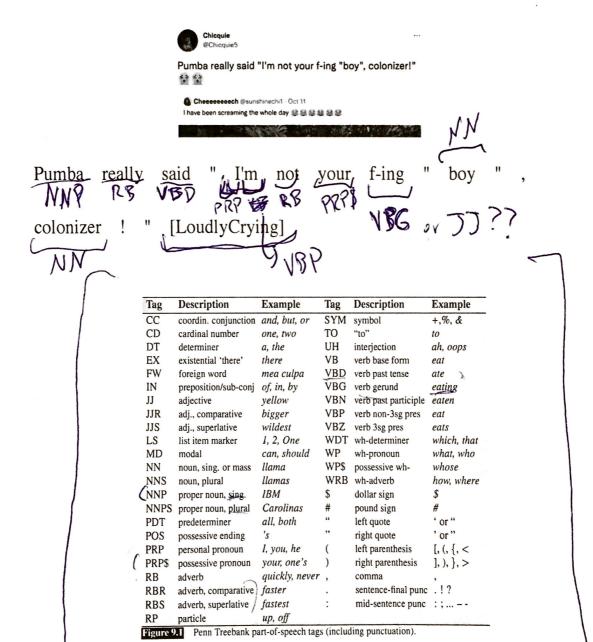- <u>Fire!</u>
  VERB

  Fire! *NOUN*

Tag these two sentences, using the PTB tagset below. You can write each tag below its word. Don't bother with punctuation, but tag all other tokens.

The New York Times

Based   on   Mr.   Trump   's   boasting   and   gaudy   lifestyle   ,

Mr,   Wallach,   imagined   he   would   soon   be   leading   impressive
NNP   *NNP*

construction   projects   around   the   globe   .

Chicquie
@Chicquie5

Pumba really said "I'm not your f-ing "boy", colonizer!"

Cheeeeeeech @sunshinechi1 · Oct 11
I have been screaming the whole day

Pumba   really   said   "   I'm   not   your   f-ing   "   boy   "   ,
NNP   *RB*   *VBD*   *PRP* *RB*   *PRP$*   *VBG* or *JJ* ??   *NN*

colonizer   !   "   [LoudlyCrying]
NN   *VBP*

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	*and, but, or*	SYM	symbol	*+,%, &*
CD	cardinal number	*one, two*	TO	"to"	*to*
DT	determiner	*a, the*	UH	interjection	*ah, oops*
EX	existential 'there'	*there*	VB	verb base form	*eat*
FW	foreign word	*mea culpa*	VBD	verb past tense	*ate*
IN	preposition/sub-conj	*of, in, by*	VBG	verb gerund	*eating*
JJ	adjective	*yellow*	VBN	verb past participle	*eaten*
JJR	adj., comparative	*bigger*	VBP	verb non-3sg pres	*eat*
JJS	adj., superlative	*wildest*	VBZ	verb 3sg pres	*eats*
LS	list item marker	*1, 2, One*	WDT	wh-determiner	*which, that*
MD	modal	*can, should*	WP	wh-pronoun	*what, who*
NN	noun, sing. or mass	*llama*	WP$	possessive wh-	*whose*
NNS	noun, plural	*llamas*	WRB	wh-adverb	*how, where*
NNP	proper noun, sing.	*IBM*	$	dollar sign	*$*
NNPS	proper noun, plural	*Carolinas*	#	pound sign	*#*
PDT	predeterminer	*all, both*	"	left quote	*' or "*
POS	possessive ending	*'s*	"	right quote	*' or "*
PRP	personal pronoun	*I, you, he*	(	left parenthesis	*[, (, {, <*
PRP$	possessive pronoun	*your, one's*	)	right parenthesis	*],), }, >*
RB	adverb	*quickly, never*	,	comma	*,*
RBR	adverb, comparative	*faster*	.	sentence-final punc	*. ! ?*
RBS	adverb, superlative	*fastest*	:	mid-sentence punc	*: ; ... -- -*
RP	particle	*up, off*			

Figure 9.1   Penn Treebank part-of-speech tags (including punctuation).

DT — VB   unlikely

DT — NN   likely

DT

fire

1)   2)   3)

Indep. LR can't use feats
from other tags