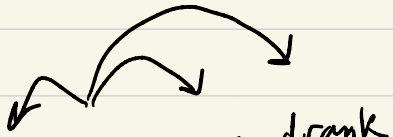
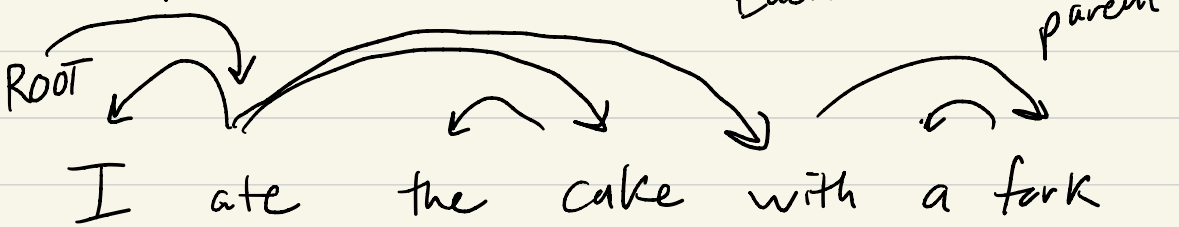
CS 378 Lecture 14

Today
- Shift-reduce parsing
- Midterm review



Recap Dependencies

- Each word has one parent



- Verbs are heads of sentences
- Verbs have nouns, prepositions as children most frequently

Advantages:
- Some attachments make more sense than in constituency
- Easier to adapt to a wide range of languages

Announcements

- A3 due
- No lecture Thursday
- Midterm: Weds 9am – Fri 5pm
  - Ask Qs via email or in private Piazza posts
  - Open book, NOT collaborative!
- Extra OHs today
  1:30–2:30
  No OHs Weds – Fri

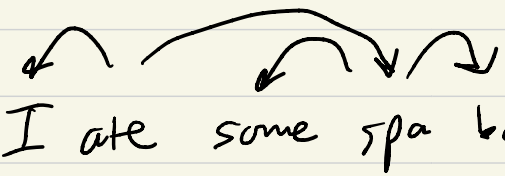
Shift-reduce parsing

- Move through a sentence word-by-word + make decisions as we go

Stack: partial parse trees

Buffer: rest of the sentence

Initial state:



Stack [ROOT]

Buffer [I ate some spaghetti bolognese]

Three ops: ① shift first word from buf → stack

reduce: left-arc
right-arc

add an arc to combine items on the stack

S [ROOT] B [I ate some spa ba]

Shift

S [ROOT I] B [ate some spa ba]

Shift

S [ROOT I ate] B [some spa ba]

Garden path: The horse raced past the barn fell.
(that was)

② Left-arc: takes top two elts of stack, makes 2nd-to-last a child of the last one, adds to stack

Left-arc

S [ROOT ate] B [some spa bo]
len 2 ↓
I

Shift Shift

S [ROOT ate some sp] B [bo]
↓
I

Left-arc (try to do these ASAP)

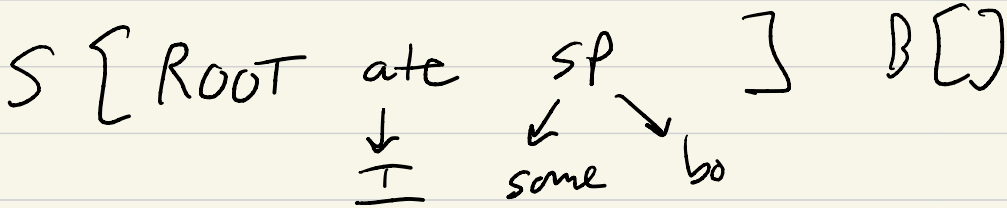
S [ROOT ate sp] B [bo]
↓ ↓
I some

Shift

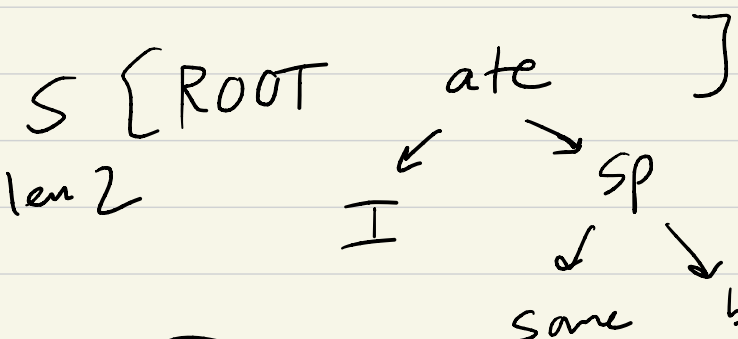
DO NOT R-A!

(3) Right-arc: takes top two elts from stack, makes last a child of 2nd-to-last, adds

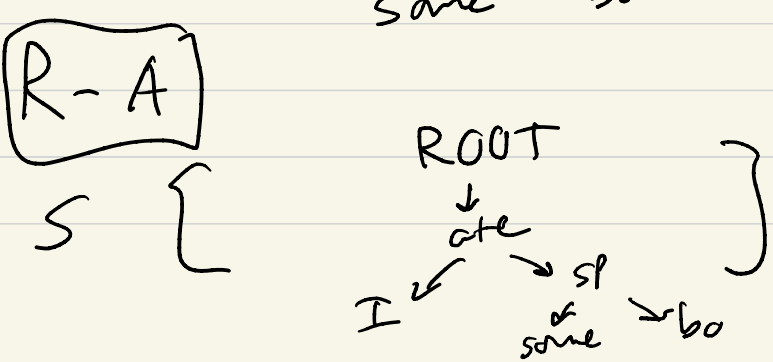
Right-arc



Right-arc



R-A



Building Shift-reduce parsers

Our parser is a classifier

Maps from state (stack, buffer)
to one of three actions

Features $f(S, B)$ ←

"Different weights" W_{SH}, W_{RA}, W_{LA} — dot product

Feats are complex!

S[.. ate] B[the ...]

Indicator [last word on stack is a verb & first Buf word is the]

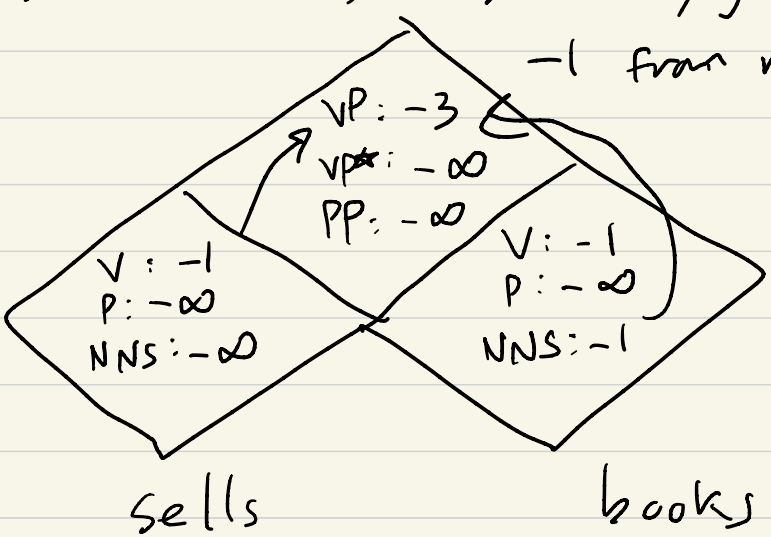
Midterm review

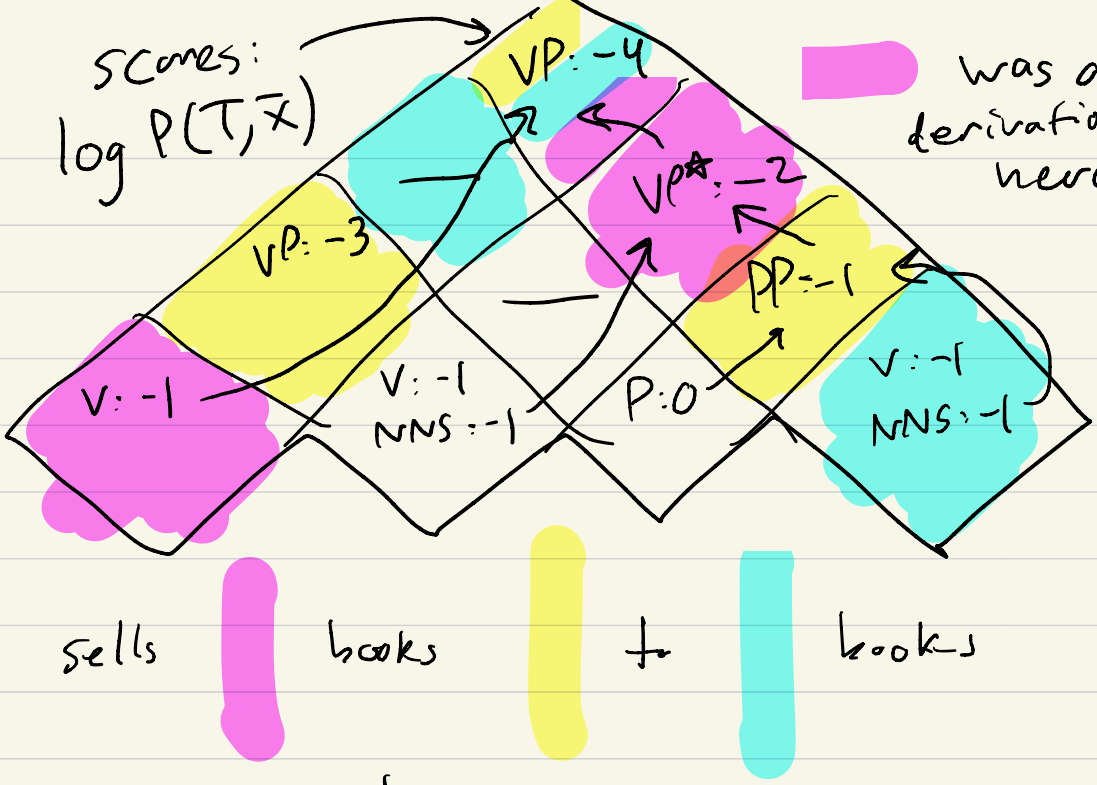
VP → V NNS PP

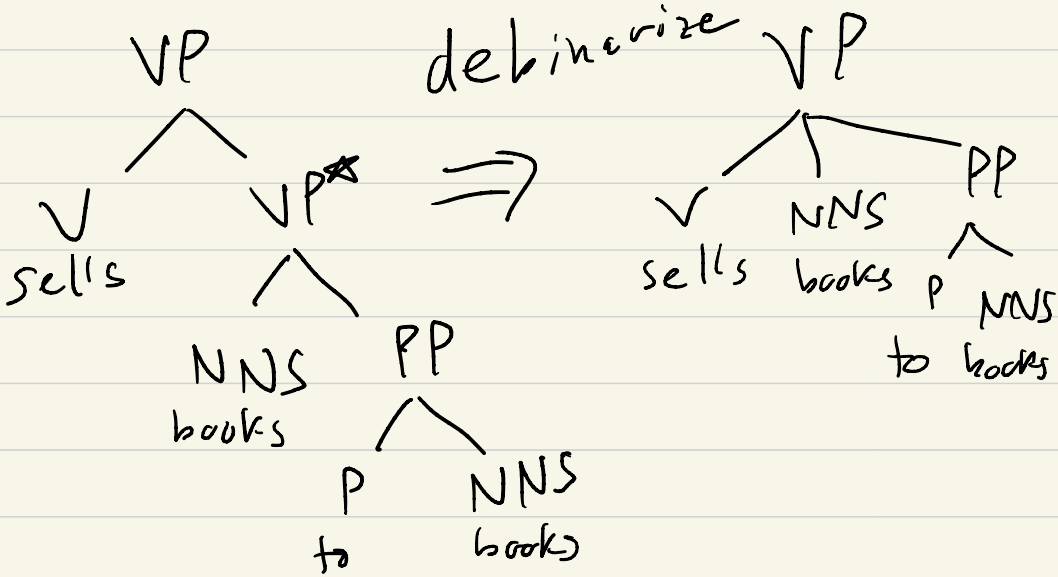
Start symbol: VP

VP → V VP*	0.5	P(rule | VP) must normalize
VP → V NNS	0.5	
VP* → NNS PP	1.0	
PP → P NNS	1.0	

rest follows the sheet, ignore Q1




scores:
log P(T,x)
VP: -4
VP*: -2
VP: -3
PP: -1
V: -1
V: -1
NNS: -1
P: 0
V: -1
NNS: -1
was our derivation here
sells
books
to
books
splits here


VP
V
sells
VP*
NNS
books
PP
P
to
NNS
books
debinarize
VP
V
sells
NNS
books
PP
P
to
NNS
books

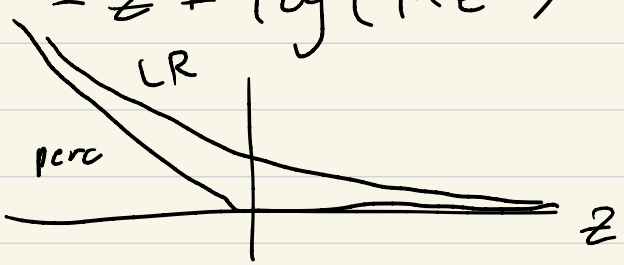
Logistic regression: loss when $y^{(i)} = +1$

$$-\log P(y=+1 \mid \bar{x})$$

$$= -\log \frac{e^{\bar{w}^T f(\bar{x})}}{1+e^{\bar{w}^T f(\bar{x})}}$$

$$= -\underbrace{\bar{w}^T f(\bar{x})}_{z} + \log\left(1 + e^{\overbrace{\bar{w}^T f(\bar{x})}^{z}}\right)$$

$$-z + \log(1+e^z)$$

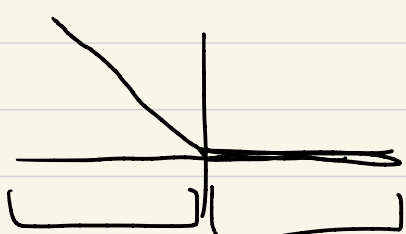


$$\frac{\partial}{\partial \bar{w}} \Big] = -f(\bar{x})$$

$$-\bar{w}^T f(\bar{x})$$

$$\begin{cases} -z & z < 0 \\ 0 & \text{else} \end{cases}$$

Perceptron:



penalize by how wrong it is

correct = 0 loss