

Mitchell Troy

NLP HW 4 Report

Part 1)

a)

There are 752 unique rules in the grammar

The top five rules are:

PUNC -> . # 346.0

TO -> to # 241.0

PP -> IN NP_NNP # 239.0

IN -> from # 218.0

PP -> IN NP # 197

b)

Top five probable rules are:

FRAG_ADJP_JJS -> Cheapest # 1.0

SBARQ* -> PP WHNP_WHNP # 1.0

X_JJ -> nonstop # 1.0

RBS -> least # 1.0

NP_CD -> one # 1.0

Note: there are many 1.0 probable rules, so your results may differ each time you run the program

Part 2)

a)

Output for first 5 lines of dev file:

Line 0

parse:

(TOP (S (NP (DT The) (NN flight)) (VP (MD should) (VP (VB be) (NP (NP* (CD eleven) (RB a.m)) (NN tomorrow)))))) (PUNC .))

Log prob: -40.8600586204

Line 1

parse:

(TOP (S (S* (S (NP_PRP I) (VP (MD would) (VP (VP* (VP* (VB like) (NP_PRP it)) (X_TO to)) (VP (VBP have) (NP (NP (DT a) (NN stop)) (PP (IN in) (NP (NNP New) (NNP York))))))) (CC and)) (S (NP_PRP I) (VP (MD would) (VP (VB like) (NP (NP (DT a) (NN flight)) (SBAR (WHNP_WDT that) (S_VP (VBZ serves) (ADVP_RB <unk>))))))) (PUNC .))

Log prob: -84.9602620899

Line 2

parse:

(TOP (SBARQ (WHNP (WHNP_WDT Which) (PP (IN of) (NP_DT these))) (SQ_VP (VBP serve) (NP_NN dinner))) (PUNC ?))

Log prob: -20.2054617111

Line 3

parse:

(TOP (SBARQ (WHNP (WDT Which) (NNS ones)) (SQ_VP (VBP stop) (PP (IN in) (NP_NNP Nashville)))) (PUNC ?))

Log prob: -23.540353035

Line 4

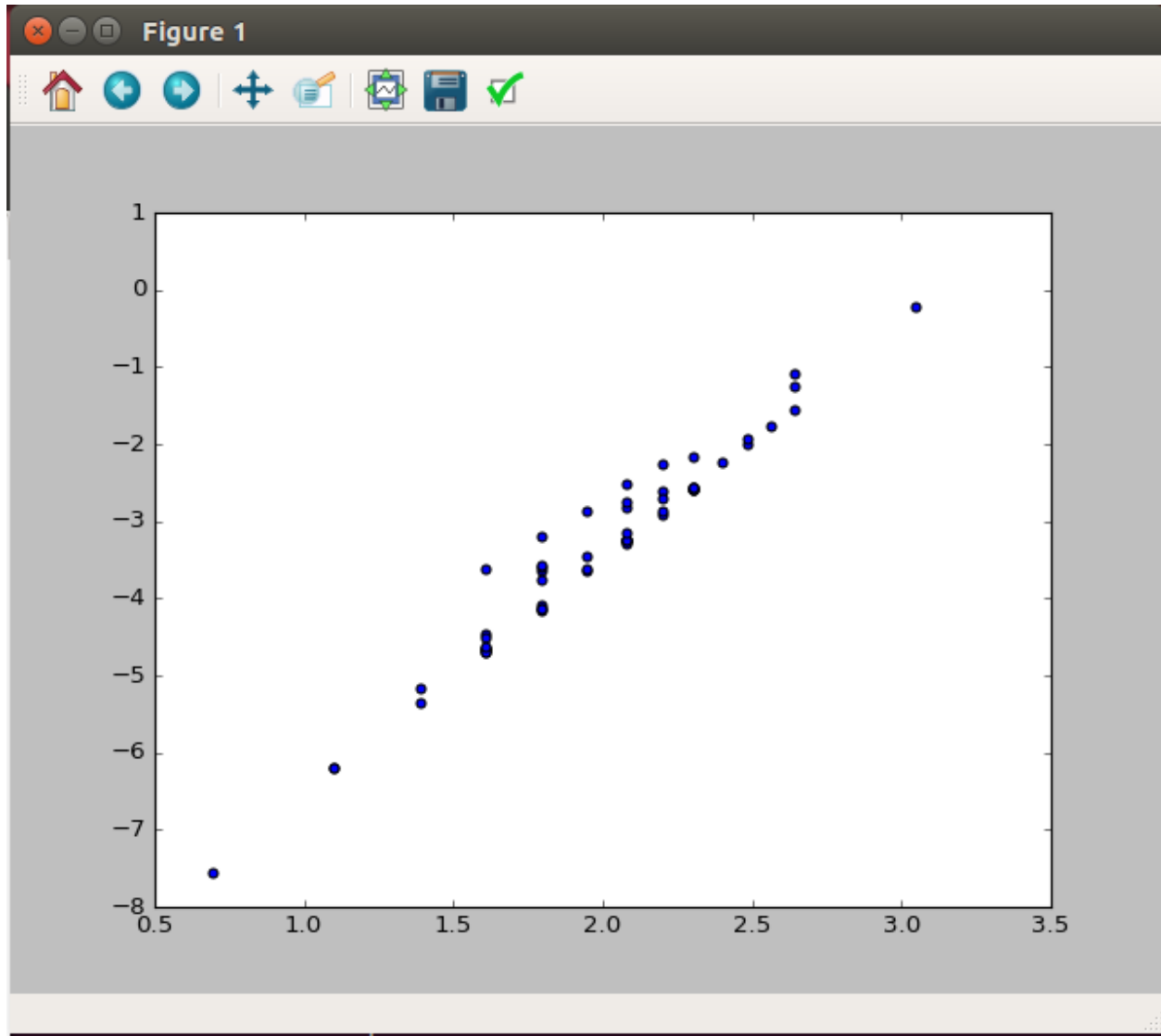
parse:

(TOP (SQ (VBP <unk>) (SQ* (NP_NP_EX there) (SQ* (NP (DT any) (NNS flights)) (VP (VBG arriving) (PP (IN after) (NP (CD eleven) (RB a.m)))))) (PUNC ?))

Log prob: -37.2789349196

b)

Log-log plot, sentence length = x-axis, parse time = y-axis



The best fit for this equation = $3.037 \log(x) - 9.396$

Note: found weights using np.polyfit function with polynomial value of 1 (to fit the equation $y = c \log(x) + d$)

Yes, my equation is very close to fitting the $y = cx^3$ equation

– it fits the precise equation $y = 8.3 \times 10^{-5} x^{3.0307}$

c)

dev.parses.post 435 brackets

dev.trees 474 brackets

matching 400 brackets

precision 0.9195402298850575

recall 0.8438818565400844

F1 0.8800880088008801