

CPSC 4310/5310/7310 – Spring 2020

Natural Language Processing (NLP)

Sample Solutions to Assignment 1 [60 points]

Due on January 24th, 2020

The problems are adopted from the textbook.

1. [5 points]

I saw a man on a hill with a telescope.

can be interpreted as:

- (a) There is a man on a hill, and I am watching him with my telescope.
- (b) There is a man on a hill, who I am seeing, and he has a telescope.
- (c) There is a man, and he is on a hill that also has a telescope on it.
- (d) I am on a hill, and I saw a man using a telescope.
- (e) There is a man on a hill, and I am sawing him with a telescope.

2. [10 points]

- (a) $[a - zA - Z]^+$
- (b) $[a - z]^* e$
- (c) $[a - z]^* ation$
- (d) $(2[5 - 9]|3[0 - 4])[0 - 9]|350$
- (e) $([0 - 9]|([1 - 9][0 - 9])|(\backslash \backslash .[0 - 9]\{1, 2\}))\%$
- (f) $([a - zA - Z]^+)\backslash s + \backslash 1$

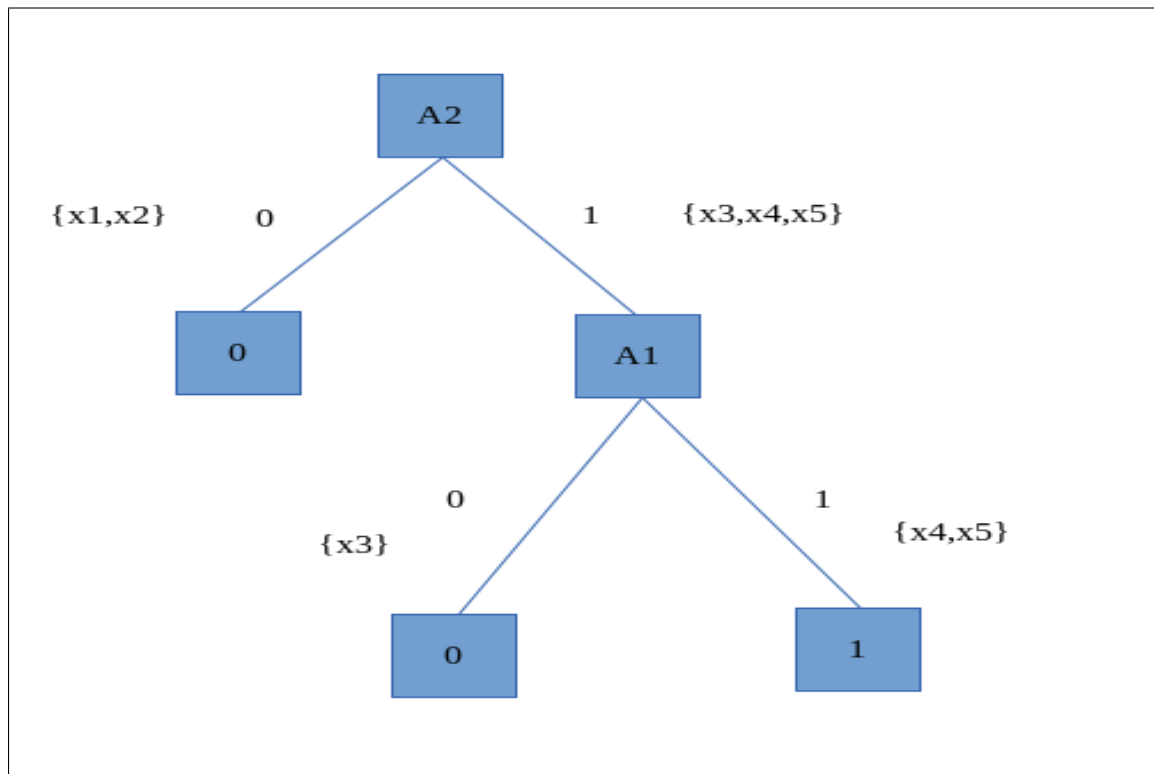
3. [5 points]

$$H(\text{Output}) = -2/5 \log_2(2/5) - 3/5 \log_2(3/5) = .53 + .44 = .97 = E(2/5, 3/5)$$

$$IG(\text{Output}, A_1) = E(2/5, 3/5) - 4/5 E(2/4, 2/4) - 1/5 E(0/1, 1/1) = .97 - .8 = .17$$

$$IG(\text{Output}, A_2) = E(2/5, 3/5) - 2/5 E(0/2, 2/2) - 3/5 E(2/3, 1/3) = .97 - .55 = .42$$

$$IG(\text{Output}, A_3) = E(2/5, 3/5) - 2/5 E(1/2, 1/2) - 3/5 E(1/3, 2/3) = .97 - .4 - .55 = .02$$



4. [10 points]

e	4	d 3	ids 4	i 3	ids 4	s 3
s	3	d 2	ids 3	i 2	ids 3	ids 4
a	2	d 1	ids 2	s 1	i 2	i 3
c	1	s 0	i 1	i 2	i 3	i 4
#	0	1	2	3	4	5
#	c	r	a	n	e	

e	4	d 3	d 2	ids 3	s 2
s	3	d 2	d 1	ids 2	ids 3
a	2	d 1	s 0	i 1	i 2
c	1	s 0	i 1	i 2	i 3
#	0	1	2	3	4
#	c	a	r	e	

So, care is closer to case than crane.