Final Exam Solutions - Spring 2021

Q1.1

TAGs model cross serial dependencies.
TAGs include an adjunction operator that makes them more powerful than CFG.

Q1.2

Bird (as a noun synset)

Q1.3

Consider the sentence *Fruit flies like a banana*.

a. This sentence is ambiguous. Paraphrase the two possible interpretations

1. Fruit flies (insect) enjoy/have a preference for a banana.
2. Fruit (food) flies (verb) in the way a banana flies.

b. Consider the following CCG parses of the sentence, which have some of their labels removed (indicated with [ ]). For each parse, fill in the missing labels.

Parse 1:

1. ((S\NP)\(S\NP))/NP
2. NP
3. (S\NP)\(S\NP)

Parse 2:

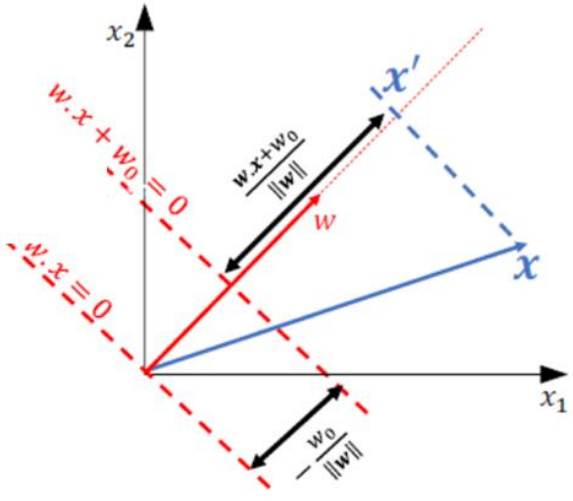
4. (S\NP)/NP
5. NP
6. S\NP

c. For each parse, indicate which of the two meanings the given parse corresponds to.

Parse 1 = Fruit (food) flies (verb) in the way a banana flies.
Parse 2 = Fruit flies (insect) enjoy/have a preference for a banana.

Q1.4

Explain the meaning of x_1, x_2, x, x', w, w_0, and ||w|| in the image below.
x1, x2 are axes
x' is the projection of x onto w
x is a vector in R2
w^tx =0 defines a line passing through the origin and orthogonal to w.
wo shifts the line wx along w.



||w|| is the length of the vector w

Q1.5

Fill in the missing semantic types:

S	-> NP VP	{VP.Sem(NP.Sem)}	t
VP	-> V NP	{V.Sem(NP.Sem)}	<e,t>
NP	-> N	{N.Sem}	e
V	-> likes	{λ x,y likes(x,y)	<e,<e,t>>
N	-> Javier	{Javier}	e
N	-> pizza	{pizza}	e

In 2021 exam: redacted 2, 3, 4, 6 so solution is <e,t>, e, <e,<e,t>>, e

Q1.6

Given the word embeddings (V_w) for the words w in {Paris, London, France, English, French, Euro, UK, etc.}, what is the value of ($V_{London} - V_{Paris} + V_{Euro}$)?

V_{pound}

Q1.7

Which of the following words start(s) with a voiced consonant?
gate, deer

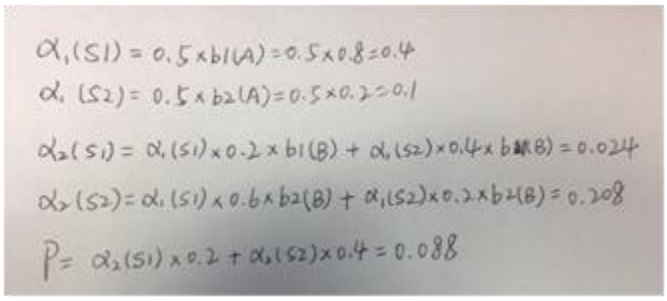
Part 2: Short Answer Questions with Numerical Answers
6 items * 4 points = 24 points

Q2.1

Using a chart, compute the edit distance between APPLE and PIE. Assume equal costs for the three basic edit operations.

3

Q2.2



Q2.3

One possible PCFG is:
S -> A p=1
A -> the p=.4
A -> A the p=.6
Any sequence of n the's requires 1x the first rule, n-1x the last rule, and 1x the middle rule.

Q2.4

P(A|B)		**A = "dog"**	
		Yes	No
B = NN	Yes	0.1	0.9
	No	0.01	0.99

Suppose *p(NN)* = 0.4. Use the Bayes rule to write the expression for *p (NN| "dog")*. Write it first as a function of *p(dog)* and then, also give the actual probability value.

p(NN|dog) = [p (dog|NN) * p(NN)] / [p(dog)]
= 0.1 * 0.4 / p(dog)
= 0.04 / p(dog)

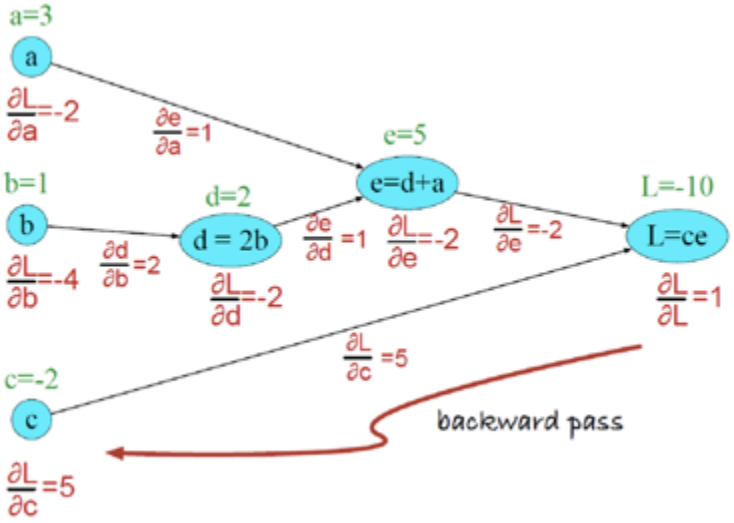
p(NN|dog) / p(~NN|dog) = X
p(NN|dog) + p(~NN|dog) = 1

Solve for p(NN|dog)

Answer should be 0.869

Q2.5

Compute all the missing partial derivatives.



Q2.6

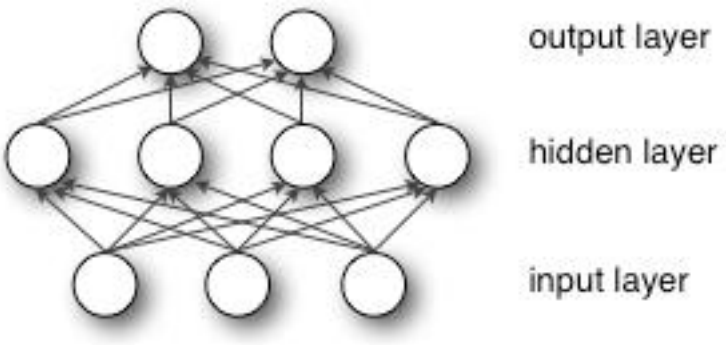
A corpus contains 1,000,000 tokens, including 6,000 instances of the word "the". The next four most frequent words are "of", "and", "to", and "a". What are the expected counts for each of these four words in the corpus?

6000/2=3000, 6000/3=2000, 6000/4=1500, 6000/5=1200

Part 3: Neural network sentiment classification
5 points

Q3

We use a 3-layer neural network for sentiment classification of words. The architecture of the neural network is shown in the picture. Words are represented by 3-dimensional embeddings as inputs, and the network outputs a probability distribution over the positive and negative classes as a 2-dimensional vector.



The hidden layer and the output layer use Rectified Linear Units (ReLU) as the activation function.

$$ReLU(t) = \begin{cases} t, if\ t \geq 0 \\ 0, otherwise \end{cases}$$

Suppose your parameters are

$$W_1 = \begin{bmatrix} -1 & 2 & 0 \\ 3 & 4 & 2 \\ 1 & -2 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$b_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$W_2 = \begin{bmatrix} 2 & -1/3 & 0 & 2 \\ 1 & 0 & 1 & 4 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

1) Suppose the input word has an embedding of:

$$x = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$$

Calculate the activation of the hidden layer.

$$h_1 = ReLU(W_1 x + b_1) = ReLU\left(\begin{bmatrix} -2 \\ 9 \\ 3 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 9 \\ 3 \\ 0 \end{bmatrix}$$

2) Following (1), what is the probability distribution after we apply softmax over the output layer activations?

$$p = softmax(ReLU(W_2 h_1 + b_2)) = softmax\left(ReLU\left(\begin{bmatrix} -3 \\ 3 \end{bmatrix}\right)\right) = softmax\left(\begin{bmatrix} 0 \\ 3 \end{bmatrix}\right)$$
$$= \begin{bmatrix} 1/(1+e^3) \\ e^3/(1+e^3) \end{bmatrix} = \begin{bmatrix} 0.0474 \\ 0.9526 \end{bmatrix}$$

We use cross-entropy loss as our objective function defined as:

$$L = -\frac{1}{N}\sum_{n=1}^{N}[y_n \log p_n + (1 - y_n)\log(1 - p_n)]$$

where N is the batch size, and we label positive words with y = 1, and negative words with y = 0.

3) Now suppose the current mini-batch contains four words: {good, bad, excellent, poor}. The network outputs a probability distribution as the following:

	good	bad	excellent	poor
P (y=1)	0.7	0.3	0.6	0.2

What is the value of the objective function for the current mini-batch?

$$L = -\frac{1}{4}(\log 0.7 + \log(1 - 0.3) + \log 0.6 + \log(1 - 0.2)) = 0.3618$$

Part 3: Multiple Choice Questions
25 items * 2 points = 50 points

Q4.1

The following sentences show examples of what linguistic phenomenon?

I had a coffee this morning (meaning "I had one cup of coffee")
I tried two wines last night (meaning "I had two types of wine")
I had fish for dinner (meaning "I had some fish", not "I had a fish")

a. **type coercion**
b. selectional restriction
c. non-projectivity
d. backoff

Q4.2

Consider the tweet below. It is funny, partially because it involves an ambiguity. What type of ambiguity is it?

a. part of speech
b. referential
c. morphological
d. syntactic

Q4.3

What is the formula for tanh(x) as a function of the sigmoid function s(x).

a. $\tanh(x)=2s(2x)-1$
b. $\tanh(x)=s(2x)$
c. $\tanh(x)=s(x)$
d. $\tanh(x)=2s(x)-1$
e. $\tanh(x)=2s(2x)$

Q4.4

What is this formula used for?

$$d = \arg\max_i \frac{(x_b - x_a + x_c)^T x_i}{||x_b - x_a + x_c||}$$

a. continuous bag of words (CBOW)
b. wordnet-based semantic similarity
c. dimensionality reduction
d. word analogy computation
e. none of the above

Q4.5

Which of these languages is closest to English from a historical, evolutionary perspective?

a. Turkish
b. Norwegian
c. Russian
d. Korean
e. Spanish

Q4.6

A dependency tree for a sentence with N words includes this many dependencies:

a. N/2
b. N-1
c. N
d. 2N
e. N(N-1)/2

Q4.7

If f(x) is coffee beans and f'(x) is ground coffee, what is f''(x)?

a. cup of water
b. cup of green tea
c. cup of black coffee
d. cup of venti non-fat ice caramel macchiato with extra foam, to go

Q4.8

Which of the following areas of linguistics deals with the derivation of the word "computer" from the word "compute"?

(a) Inflectional morphology.
(b) Derivational morphology.
(c) Lexical semantics.
(d) Compositional semantics.
(e) None of the above.

Q4.9

The cosine similarity between the vectors (1,2,0) and (1,1,2) is:

(a) 0.5
(b) 0.32
(c) 1
(d) **0.55**
(e) 0.80

Q4.10

A collection includes 1,000,000 documents. The word "saline" appears in 1,000 of these documents: in 200 of them it appears once in each document, in the next 500 of them it appears twice in each document, and in the final 300 of them it appears nine times in each document. What is the IDF (inverse document frequency) of the word "saline", rounded to the nearest integer?

(a) **3**
(b) 11
(c) 5
(d) 9
(e) 7

Q4.11

How would you represent the sentence "Exactly one student passed the test" in First Order Logic (FOL)?

(a) **∃x: student(x) ∧ passed (x, test) ∧ [∀y: (student (y) ∧ passed(y, test)) ==> x=y]**
(b) ∀x : student (x) ==> ¬ passed (x, test)
(c) ∃x: student (x) ∧ ¬ passed (x, test)
(d) ¬ [∃x : student (x) ∧ ¬ passed (x, test)]
(e) ∃x : student (x) ∧ passed (x, test) ∧ [∃y : student (y) ∧ passed (y, test) ∧ x = y]

Q4.12

Consider the following two sequences of parts of speech (each sequence corresponds to a sentence).

DT JJ NN PRP VBP TO VB DT NN IN VB IN DT NNS
DT NN IN DT NN NN RB VBZ PRP VB RP NNS

What is the maximum likelihood estimate (MLE) for the probability of bigram "NN IN"?

(a) 0
(b) .6
(c) .25
(d) **.4**
(e) .1

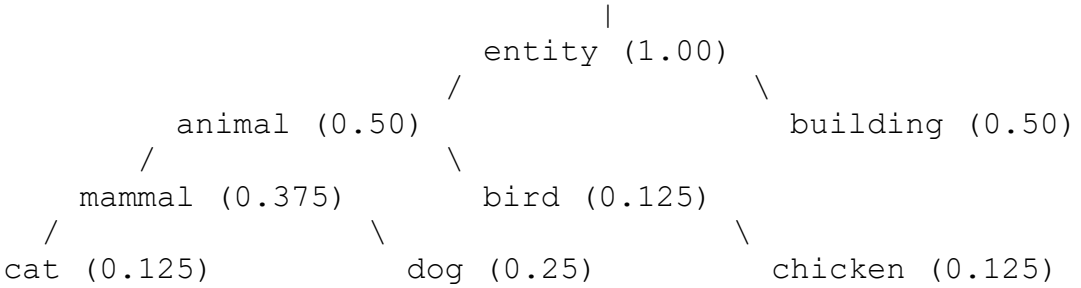
Q4.13

What is the complex type for a preposition in CCG?

(a) (NP\NP)/NP
(b) (S\NP)/NP
(c) NP/N
(d) NP\NP
(e) S/NP

Q4.14

Consider the following segment of a Wordnet-like ontology, augmented with subtree probabilities. What is the "Lin" similarity between chicken and dog?



(a) 0.125
(b) 0.200
(c) 0.250
(d) 0.400

(e) 0.500

Q4.15

What is "acerola"?

(a) a fast train connecting New York to Boston
(b) a fruit with soft pulp
(c) a roughly straight-line configuration of three or more celestial bodies in a gravitational system
(d) a deep learning library
(e) a Pokemon character

Q4.16

According to the rules of ITG (Inversion Transduction Grammar), how many reorderings are allowed for the production

```
NP -> ART CARD JJ NN?
```

(a) 1
(b) 2
(c) 4
(d) 16
(e) 24

Q4.17

True or False: Given the feature structures FS1 and FS2,

FS1

```
[ agr  = [ number = 'singular' ] ]
[        [ person = 1          ] ]
[                                ]
[ type = 'NP'                    ]
```

FS2

```
[ agr  = [ number = ?n ] ]
[                        ]
[ subj = [ number = ?n ] ]
```

the output of their unification Unify (FS1, FS2) is correctly shown below.

FS3

```
[ agr  = [ number = 'singular' ] ]
[        [ person = 1          ] ]
[                                 ]
[ subj = [ number = 'singular' ] ]
[                                 ]
[ type = 'NP'                     ]
```

(a) TRUE
(b) FALSE
(c) no idea

Q4.18

IBM Models 1, 2, and 3 are used for:

(a) statistical machine translation
(b) neural machine translation
(c) constituent parsing
(d) dependency parsing
(e) none of the above

Q4.19

In order to prevent the possibility of an artificially high machine translation score, BLEU includes the following component:

(a) counting crossing brackets
(b) finetuning
(c) “forget” gates
(d) brevity penalty
(e) subcategorization

Q4.20

What grammatical formalism is specifically designed to handle sentences like this one: “I bought a hat for my son and a book for my daughter”?

(a) Head Driven Phrase Structure Grammar (HPSG)
(b) Dependency Grammar
(c) Regular Grammar
(d) Context Sensitive Grammar (CSG)
(e) Combinatory Categorial Grammar (CCG)

Q4.21

Given the sentence “When you get home, I will have fixed the sink”, in what order do the utterance (U), reference point (R), and event (E) occur? The “<” symbol below means “precedes”.

(a) E < R < U
(b) U < E < R
(c) U < (R,E)
(d) (U,R) < E
(e) E < U < R

Q4.22

Which of the following techniques is *not* used by BERT:

(a) bidirectional encoding
(b) attention
(c) parallel corpora
(d) stacked encoders
(e) position embeddings

Q4.23

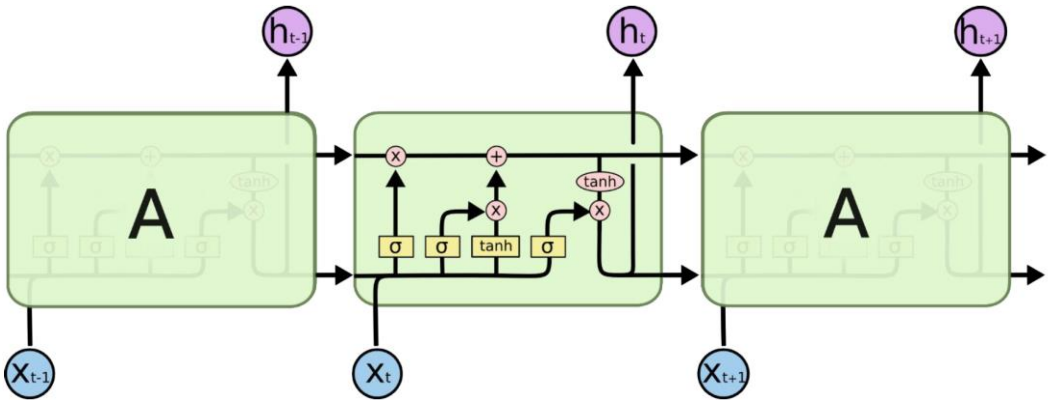
What type of neural network uses the equations below?

$$h_t = \sigma(W_h h_{t-1} + W_x x_t)$$
$$y_t = softmax(W_y h_t)$$

(a) recurrent neural network
(b) recursive neural tensor network
(c) convolutional network
(d) gated recurrent unit
(e)long short term memory network

Q4.24

What type of neural network is represented here (image from Chris Olah)?



(a) recurrent neural network
(b) recursive neural tensor network
(c) convolutional network
(d) gated recurrent unit
(e)long short term memory network

Q4.25

In the sentence “She is eating a red apple”, which of the following pairs of words would exhibit the *highest* attention scores?

(a) she, red
(b) is, a
(c) eating, red
(d) she, apple
(e) eating, apple

Part 4: Bonus Question
5 points

Q5 -- Question 961 (BONUS QUESTION)

1. <1, 4, 4, 0, 1, 2, 1, 1, 1, 1, 0, 0, 0, 0, 0>

2.

a)	*antismartnessesquely*		*X4*
b)	*aunt*	*X5*	
c)	*cats*	*X1*	
d)	*meow*	*X3*	
e)	*the*	*X2*	