

Test Information

Description After you complete the assignment, you may get a mark (probably 0), it is a dummy result and not the real result because the true answers have not been input to the systems yet. The real result will be published 2 weeks or more after the assignment deadline.

Instructions Hard deadline: late submission is not allowed.

Please click on the "save and submit" button at the end of the page after you completed the assignment.

If your answer is not an integer, give it to the nearest 3 decimal places.

If your answer is an integer, add decimal point has no harm. E.g. 3=3.0=3.00=3.000 etc.

Please complete the assignment before the deadline. Multiple attempts are allowed.

Multiple Attempts This test allows multiple attempts.

Force Completion This test can be saved and resumed later.

Your answers are saved automatically.

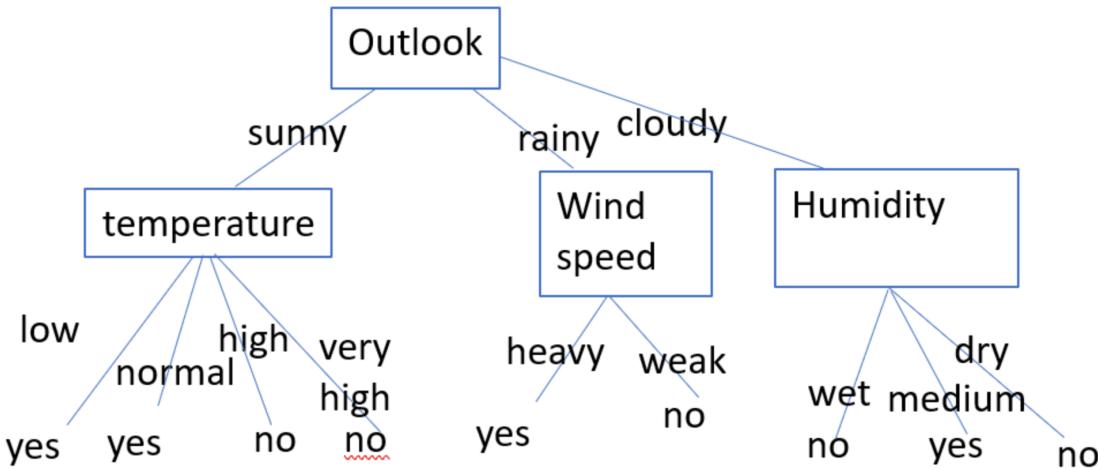
Question Completion Status:

QUESTION 1

10 points

Save Answer

In the diagram, the number of nodes in this tree is N, the number of leaves is E, find N+E.



QUESTION 2

10 points

Save Answer

Attributes				classes
Feature1	Feature2	Feature3	Feature4	class
A	1	X	L	Decision YES
A	2	Y	H	Decision YES
B	2	X	L	Decision NO
B	1	X	H	Decision NO
C	1	X	L	Decision NO
C	2	Y	L	Decision NO
A	1	Y	H	Decision NO
B	1	Y	H	Decision NO
B	1	X	L	Decision YES
A	2	Y	H	Decision NO

As part of the steps for developing a decision tree, we need to find the entropy values of the nodes. Please find the entropy of the parent node of the decision tree using the given table .

QUESTION 3

10 points

Save Answer

As part of the steps for developing a decision tree, we need to find the Gini index of the nodes. Please find the Gini index of the parent node of the decision tree using the given table.

Attributes				classes
Feature1	Feature2	Feature3	Feature4	class
A	1	X	L	Decision 1
A	2	Y	H	Decision 1
B	2	X	L	Decision 2
B	1	X	H	Decision 2
C	1	X	L	Decision 2
C	2	Y	L	Decision 1
A	1	Y	H	Decision 2
B	1	Y	H	Decision 3
B	1	X	L	Decision 2
A	2	Y	H	Decision 1

QUESTION 4

15 points

Save Answer

The weather and the decision of wearing thick clothes are shown in the given table.

The value of Information gain is A by entropy if Humidity is the root of the tree.

The value of Information gain is B by entropy if Temperature is the root of the tree.

The value of Information gain is C by entropy if Wind is the root of the tree.

Give you answer as A+B+C.

Samples	Weather			Decision:
	Feature: Humidity	Feature: Temperature	Feature: Wind	Wear thick cloth
1	Dry	Cold	High	Yes
2	Dry	Cold	Low	No
3	Medium	Average	High	Yes
4	Medium	Average	Low	No
5	Medium	Hot	High	No
6	Medium	Hot	Low	No
7	Wet	Cold	Low	Yes
8	Wet	Average	Low	No
9	Wet	Hot	High	Yes

QUESTION 5

10 points

Save Answer

Given the following corpus of sentences

- Sentence s1 : Upon the road I met seven wives
- Sentence s2 : Every wife had seven sacks
- Sentence s3 : Every sack had seven cats
- Sentence s4 : Every cat had seven kits

In this question, the corpus should be preprocessed by the stemming algorithm, i.e. catty and cats are the same word as cat, etc.

BOW (Bag Of Words) cosine similarity measure (cosine_BOW_x) can be used to determine the similarity between sentences.

Write the highest cosine_BOW_x value among two different sentences in the above corpus.

QUESTION 6

15 points

Save Answer

Given the following corpus of sentences

- Sentence s1 : Upon the road I met seven wives
- Sentence s2 : Every wife had seven sacks
- Sentence s3 : Every sack had seven cats
- Sentence s4 : Every cat had eight kits

TF-IDF (Term Frequency-Inverse Document Frequency) can be used to measure the cosine similarity difference between sentences.

In this question, the corpus should be preprocessed by the stemming algorithm, i.e. catty and cats are the same word as cat, etc.

write the TF-IDF cosine similarity difference between sentence s2 and s3.

Given a 3-neuron recurrent neural network (RNN) as in the diagram attached.
The weights used in the recurrent neural network (RNN) are whx, whh.

%assume whx,whh,why are initialized at t=1 as

whx=[0.08 0.64 0.77 0.38

0.10 0.57 0.29 0.48

0.23 0.19 0.15 0.29];

whh =[0.21 0.22 0.23

0.31 0.44 0.26

0.41 0.54 0.36];

why=[0.47 0.97 0.83

0.39 0.28 0.65

0.84 0.29 0.33

0.51 0.32 0.12];

bias=[0.41, 0.12, 0.63]';%bias initialised

ht(:,1)=[0.3 0.2 0.1]';%assume ht has these values initially at t=1

The input at time t=1 is $[X_1, X_2, X_3, X_4] = [1, 0, 0, 0]$

Find output h of the third neuron at time t=2, $h(3, t=2)$.

- Inside A : 3 neurons at time from t to t+1.

Bias=[bias(1),

bias(2), whx=

bias(3)]

h=[h1,h2,h3]'

$h_t(1)$

$h_t(2)$

$h_t(3)$

$X = [X(1), X(2), X(3), X(4)]'$,

$whx = [whx(1,1) \quad whx(1,2) \quad whx(1,3) \quad whx(1,4)]$

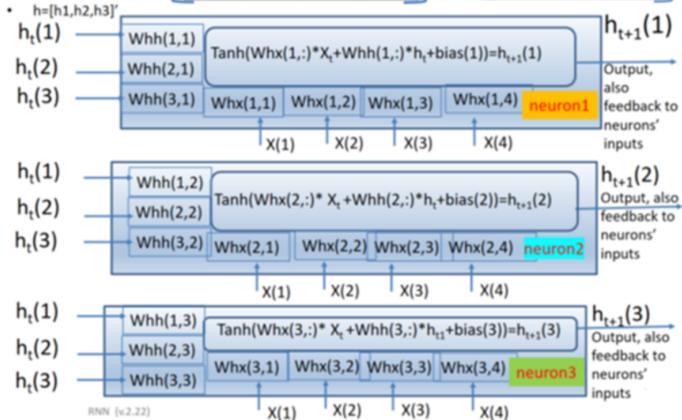
$whx = [whx(2,1) \quad whx(2,2) \quad whx(2,3) \quad whx(2,4)]$

$whx = [whx(3,1) \quad whx(3,2) \quad whx(3,3) \quad whx(3,4)]$

$whh = [whh(1,1) \quad Whh(1,2) \quad Whh(1,3)]$

$Whh(2,1) \quad Whh(2,2) \quad Whh(2,3)]$

$Whh(3,1) \quad Whh(3,2) \quad Whh(3,3)]$



whx=	<table border="1"> <tr><td>whx(1,1)</td><td>whx(1,2)</td><td>whx(1,3)</td><td>whx(1,4)</td></tr> <tr><td>whx(2,1)</td><td>whx(2,2)</td><td>whx(2,3)</td><td>whx(2,4)</td></tr> <tr><td>whx(3,1)</td><td>whx(3,2)</td><td>whx(3,3)</td><td>whx(3,4)</td></tr> </table>	whx(1,1)	whx(1,2)	whx(1,3)	whx(1,4)	whx(2,1)	whx(2,2)	whx(2,3)	whx(2,4)	whx(3,1)	whx(3,2)	whx(3,3)	whx(3,4)
whx(1,1)	whx(1,2)	whx(1,3)	whx(1,4)										
whx(2,1)	whx(2,2)	whx(2,3)	whx(2,4)										
whx(3,1)	whx(3,2)	whx(3,3)	whx(3,4)										

Input X to h output weights
(not recurrent)

whh=	<table border="1"> <tr><td>whh(1,1)</td><td>Whh(1,2)</td><td>Whh(1,3)</td></tr> <tr><td>Whh(2,1)</td><td>Whh(2,2)</td><td>Whh(2,3)</td></tr> <tr><td>Whh(3,1)</td><td>Whh(3,2)</td><td>Whh(3,3)</td></tr> </table>	whh(1,1)	Whh(1,2)	Whh(1,3)	Whh(2,1)	Whh(2,2)	Whh(2,3)	Whh(3,1)	Whh(3,2)	Whh(3,3)
whh(1,1)	Whh(1,2)	Whh(1,3)								
Whh(2,1)	Whh(2,2)	Whh(2,3)								
Whh(3,1)	Whh(3,2)	Whh(3,3)								

Current h_t to next h_{t+1} weights
(recurrent)

QUESTION 8

15 points Save Answer

The output system of a recurrent neural network (RNN) is as follows.

If $ht(1)=0.45$, $ht(2)=0.56$, $ht(3)=0.84$

$why=[0.24 \ 0.27 \ 0.35$

$0.15 \ 0.18 \ 0.21$

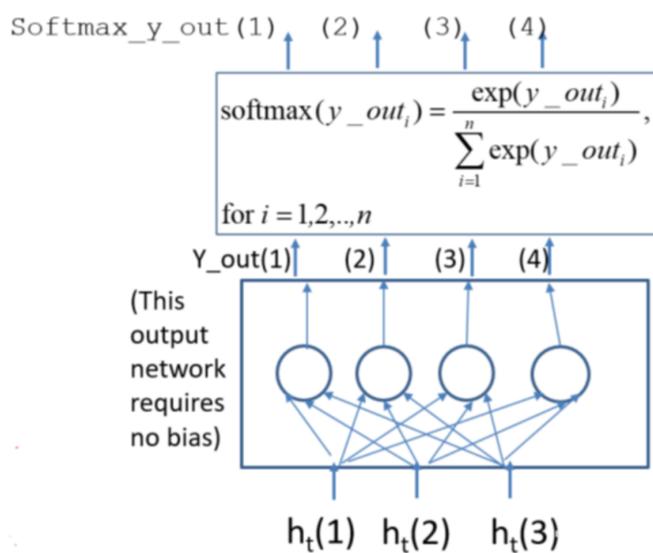
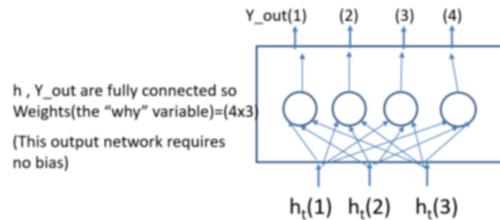
$0.25 \ 0.26 \ 0.64$

$0.86 \ 0.13 \ 0.41]$

Find $\text{Softmax}_y_{\text{out}}(3)$.

why(1,1)	why(1,2)	why(1,3)
why(2,1)	why(2,2)	why(2,3)
why(3,1)	why(3,2)	why(3,3)
why(4,1)	why(4,2)	why(4,3)

- $Y_{\text{out}}(1) = \text{why}(1,1) * ht(1) + \text{why}(1,2) * ht(2) + \text{why}(1,3) * ht(3)$
- $Y_{\text{out}}(2) = \text{why}(2,1) * ht(1) + \text{why}(2,2) * ht(2) + \text{why}(2,3) * ht(3)$
- $Y_{\text{out}}(3) = \text{why}(3,1) * ht(1) + \text{why}(3,2) * ht(2) + \text{why}(3,3) * ht(3)$
- $Y_{\text{out}}(4) = \text{why}(4,1) * ht(1) + \text{why}(4,2) * ht(2) + \text{why}(4,3) * ht(3)$

**QUESTION 9**

10 points Save Answer

For a Long short-term memory (LSTM) neural network , Input =200 neurons, Output = 100 neurons, Hidden Layers = 4, Cells in each hidden layer = 256 neurons (same for all hidden layers).

Note: for the output layer, the last hidden layer uses the sigmoid activation function to be sent to the output neurons.

The total number of weights is W, and the total number of biases is B in this neural network.

Find $W+B$.

QUESTION 10

10 points Save Answer

Multiple answers

Which of the following statement(s) is/are true?

- In a classical/vanilla autoencoder the number of input and output neurons are the same.
- In a variational autoencoder the number of input and output neurons are the same.
- In a variational autoencoder a random generator is used inside the hidden layer to generate hidden data (z) from the hidden-mean-neurons and hidden-variance-neurons.
- In a classical/vanilla autoencoder a random generator is used inside the hidden layer to generate hidden data (z) from the hidden-mean-neurons and hidden-variance-neurons.
- A variational autoencoder can be used to generate new data from a small dataset with similar probability distribution values.

QUESTION 11

10 points Save Answer

Which of the following statement(s) is/are true?

- A random process is involved in a variational autoencoder so that backpropagation can be executed directly.
- A random process is involved in a variational autoencoder so that backpropagation cannot be executed directly.
- The Re-parameterization trick in a variational autoencoder can turn a random variable into a real number.
- The Re-parameterization trick in a variational autoencoder can turn a real number into a random variable.

QUESTION 12

10 points Save Answer

Which of the following statement(s) is/are true?

- Using Sigmoid as the activation function for a neuron will cause the vanishing gradient problem.
- Using Relu as the activation function for a neuron will cause the vanishing gradient problem.
- LSTM can be used to reduce the effect of the vanishing gradient problem.
- LSTM uses more memory compared to an RNN system with the same number of input and output.

QUESTION 13

15 points Save Answer

A transformer encoder model similar to that described in the lecture notes (see the attached diagram) has the following specifications:
The system can handle a vocabulary of 50257 tokens.

Input sequence length= 2048.

Input embedded vector size = 1024.

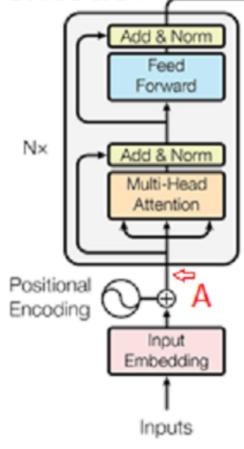
Number of attention heads= 96.

Dimension of its hidden layer = 12288.

Calculate the total number of weights used in between point A and B.

Note: Assume there is only one (linear + RELU) layer in the "Feed Forward" block in our example. (updated 2024.11.12)

Transformer encoder B ↴



QUESTION 14

10 points

Save Answer

Given learning rate =1, discount rate=0.8, if the R and the current Q tables are given as shown.

Currently the agent is at state s=2 and action a=3 is selected. An entry in the Q table will be updated, write the updated value of that entry.

State	Action					
	0	1	2	3	4	5
0	-1	-1	-1	-1	0	-1
1	-1	-1	-1	0	-1	100
2	-1	-1	-1	0	-1	-1
3	-1	0	0	-1	0	-1
4	0	-1	-1	0	-1	100
5	-1	0	-1	-1	0	100

	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	0	0	64	0	100
2	0	0	0	0	0	0
3	0	80	45	0	130	0
4	0	0	0	122	0	100
5	0	0	0	0	0	0

QUESTION 15

10 points

Save Answer

Run the code in the following link in COLAB

https://colab.research.google.com/drive/1N7HELWImK9xCYheyozVP3C_McbiRo1nb#scrollTo=dfo-brEm-STg

The following modification is needed to run the code successfully:

```
vocab = vectorizer.get_feature_names_()
should be replaced by
vocab = vectorizer.get_feature_names_out_()
```

Run the code from the beginning till the part before "Word2Vec and GloVe" to learn how to use the libraries.

You are now given 4 sentences:

```
(sentence1): The sun has long been set.
(sentence2): The stars are out by twos and threes.
(sentence3): The little birds are piping yet.
(sentence4): Among the bushes and trees.
```

Find the TF-IDF cosine similarity between sentence 2 and 4.

Hints: use the COLAB program in the above link to find the TF-IDF encoding vector. Then, you may use python (or Matlab) for the cosine similarity calculation.

QUESTION 16

10 points

Save Answer

Word embedding word2vec, programming using Anaconda,

It is suggested you install anaconda (<https://www.anaconda.com/>) and use it for this exercise. Please also read the tutorial at <https://code.google.com/archive/p/word2vec/>

```
# Download the pretrained word2vec ('GoogleNews-vectors-negative300.bin.gz') from https://drive.google.com/file/d/0B7XkCwpI5KDYNINUTTIS21eQmM/edit?resourcekey=0-wjGzDNUop6WyKtMip3g
```

Unzip 'GoogleNews-vectors-negative300.bin.gz' into a directory of your computer.

In the same directory you may use the following .py program segment to read the word2vec data.

```
#code begins #####
import numpy as np
import gensim # >pip install gensim #if needed
model = gensim.models.KeyedVectors.load_word2vec_format('GoogleNews-vectors-negative300.bin', binary=True)
```

```
rocket = model['rocket']
#Add your own code here.
#
#
```

```
#code ends #####
```

```
If
A= cosine_similarity between the word "rocket" and "plane",
B= cosine_similarity between the word "ship" and "banana".
```

Find A-B.

QUESTION 17

10 points Save Answer

Run the following steps to learn how a Q-learning program written in Python works.

1. Study how to run a python program using [anaconda\(python system\)](#). You may need libraries : gymnasium, numpy, matplotlib, pickle to successfully run the program. I.e. if you missed the library gymnasium, try to run "> pip install gymnasium" under anaconda.
2. Download https://github.com/johnnycode8/gym_solutions. Edit the program [frozen_lake_q.py](#).
3. Based on the original [frozen_lake_q.py](#), at the end of the program,
 - change the line "#run(15000)" to "run(999)".
 - Comment the following line

```
run(1000, is_training=True, render=True),  
i.e. # run(1000, is_training=True, render=True)  
Run frozen_lake_q.py
```

Observe the display and see if the man can find the exit (treasure box) or not.

4. Based on the original [frozen_lake_q.py](#), at the end of the program,
 - change the line "#run(15000)" to "run(15000)".
 - Comment the line : `run(1000, is_training=True, render=True)`,

```
i.e. # run(1000, is_training=True, render=True)  
Run frozen_lake_q.py
```

Observe the display and see if the man can find the treasure box or not.

5. Run the original [frozen_lake_q.py](#)

Observe the display and see if the man can find the treasure box or not.

If enough episodes were run, the man can find the treasure box.

No need to submit the code, just answer the following question.

- When the episode number is set to 999, the learning is not successful to solve the problem.
- When the episode number is set to 15000, the learning is successful to solve the problem.
- The Q table is stored in the file [frozen_lake8x8.png](#)
- During learning, the value of epsilon is changing from 0 to 1
- The value of "reward" in the program starts to rise from episode 12000 onward.

QUESTION 18

10 points Save Answer

Run the following steps to learn how a deep Q-learning (dql) program written in Python works.

1. Study how to run a python program using [anaconda\(python system\)](#). You may need libraries : gymnasium, numpy, matplotlib, pickle to successfully run the program. I.e. if you missed the library gymnasium, try to run "> pip install gymnasium" under anaconda.
2. Download https://github.com/johnnycode8/gym_solutions. Edit the program [frozen_lake_dql.py](#).
3. Run [frozen_lake_dql.py](#).

No need to submit the code, just answer the following question.

Which of the following statement(s) is/are true?

- The neural network parameters are stored in [frozen_lake_dql.pt](#).
- The number of neurons in the hidden layer of the neural network policy_dqn is 32.
- This DQN employs 2 neural networks of the same size.
- Replay memory is used to store neural work weights and biases.
- Copy the target network to policy network once every 10 steps.