COMP 6721 Applied Artificial Intelligence (Summer 2023)
Assignment #2 (5% of the final grade)

Submission Deadline: Tuesday June 20th, 11:59 PM (PDF on Moodle)

No need to prepare your submission using a text processor (Word, Latex). Take legible photos of your handwritten solutions, make it one PDF file.

Thank you!

Question	Subject	Points
Q1	Backpropagation	16
Q2	Convolutional Networks	13
Q3	TF-IDF and Cosine Similarity	6
Q4	Parse Tree	5
Q5	PRG MNIST CNN	20
Q6	PRG News Dataset [Bonus]	10
Q7	Compare CONV architectures [Bonus]	5

Grades: 50 + 25 Bonus = 75 pts

Please also find attached:

Item	Question
MNIST_CNN.ipynb	Question 5 MNIST CNN
Archive.zip (json file inside)	Question 6 News Dataset
News_Dataset.ipnyb	Question 6 News Dataset

Question 1) Backpropagation [15 pts]

Consider the following neural network. It consists of 2 input nodes, 2 hidden nodes, and 1 output node. All nodes in the hidden layer use the Relu activation function, and the output layer uses a Sigmoid for classification.
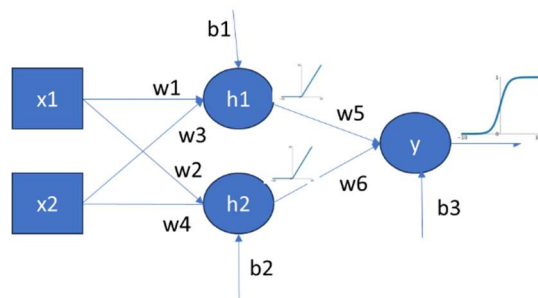
Parameter	Current value
w1	0.1
w2	0.2
w3	0.3
w4	0.4
w5	0.5
w6	0.6
b1	0.1
b2	0.2
b3	0.3

Since it is a binary classification, we have decided to use a **cross entropy loss function**. To mitigate overfitting, we have also added a regularization term equal to $\frac{1}{2}||w||^2$.

Therefore, for one example, the loss function would be

$$L = -[T \ln O + (1 - T) \ln(1 - O] + \frac{1}{2} \sum wi^2$$

Where w is the vector of all weights in the network.

a) [3pts] Calculate the output of y if the network is fed x = (1, 0) as input.
b) [2pts] Calculate the loss for the case above.
c) [9pts] Assume that the expected output for the input x = (1, 0) is supposed to be 1. Calculate the updated weights for w1, w5, and w6, after the backpropagation of the error for this sample.

Assume that the learning rate η is equal to 0.1.

Hint: (Slide 15 Optimization) The update rule of the weights will be $W_{new} = W_{old} - \eta \frac{\delta L}{\delta w}$

d) [2pts] How does the updated value of w5 change if we use Momentum with the following parameters: $\alpha = 0.9 \; and \; current \; velocity = 0.5$

Question 2) [14p] Convolutional Networks

Consider the following convolutional neural network (CNN) that predicts the class of images between two available classes (binary classification). Color RGB Images with the size of 12x12 train this network. In the first convolutional layer, we do not want any down sampling to occur, so we use a technique to make sure that the width and height of the output feature map is the same as the image size.

Fully Connected (FC)
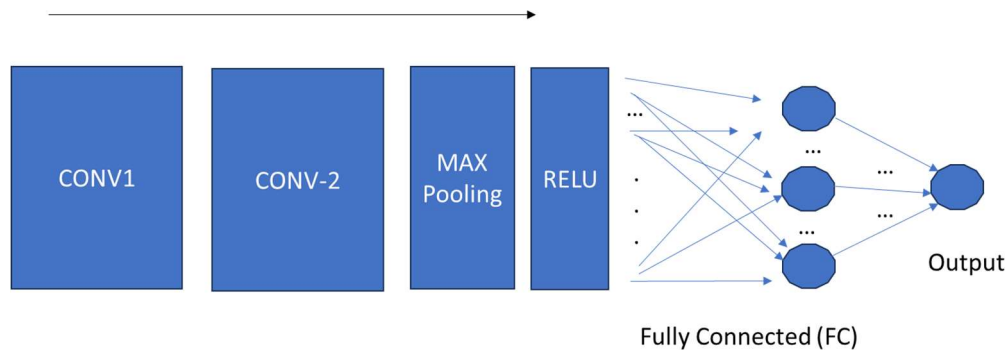
Images	First Convolutional layer:	Second Convolutional layer:	Max-pooling	FC Layer	Output
Image size: 12x12 RGB color pictures	filter size: 3x3 16 filters stride=1	filter size: 3x3 8 filters stride=1	2x2 "filters" Stride=2 (no overlap)	Relu activation	Binary classification

(a) [2p] How do you make sure that in the first CONV layer, the width and height of the input volume are preserved (so input and output will have the same width and height dimensions)?

(b) [3p] How many <u>neurons</u> in the first convolutional layer are required? (Hint: We have 3D feature maps) [Justify]

(c) [3p] How many <u>weights</u> in the first convolutional layer do we need to learn? (Ignore the bias terms of the neurons) [Justify]

(d) [2p] How many <u>neurons</u> in the second convolutional layer are required [Justify]

(e) [2p] How many <u>weights</u> in the second convolutional layer do we need to learn? (Ignore the bias terms of the neurons) [Justify]

(f) [2p] How many <u>features</u> does the whole convolutional layers provide for the FC part (i.e., output volume of the Relu layer after the max-pooling) [Justify]?

Question 3 [5pts] TF-IDF and Cosine Similarity

Compute the cosine similarity score between the query "best AI student" and the document "AI student is an intelligent student". Hint: We should find the TF-ITF weights, then we can find the cosine similarity. Please write all the steps. Use the logarithmic approach discussed in the class.

You can consider "an" and "is" as stop words. The total number of documents in the system is 80,000, and the number of documents in the system containing each word is as follows:

Word	No. de Documents
AI	10,000
Best	50,000
Intelligent	5,000
Student	1,000

Question 4) [5pts] Parse Tree

Consider the following context-free grammar. Generate all possible parse trees for the following sentence. Justify your answer. *"Book a hotel room".*

S → NP VP
S → Aux NP VP
S → VP
NP → Det NOM
NOM → Noun
NOM → Noun NOM
VP → Verb
VP → Verb NP

| |
|---|
| Det → this \| that \| the \| a |
| Noun → hotel\| flight \| book\| room |
| Verb → book \| take \| read |
| Aux → does |

Programming Question

Question 5 [20 pts] MNIST CNN

In the code (notebook) entitled mnist_cnn, you should analyze the code and answer the following questions. For some parts, you should complete the code or modify it and run it. Then, you attach your output to your submission.

For all the plots that you attach, you should add your name and student ID in the title of the plot (Example attached)

Batch Size = 128 (Arash Azarfar 12345678)
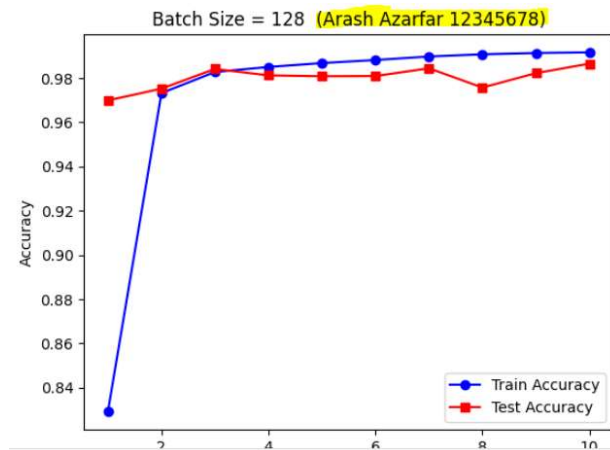
a) [4pts] What is the structure of the used CNN model?
b) [1pt] Which activation function(s) have been used in this CNN?
c) [2pts] What are these bias and weight parameters printed here? What is their role? (Check slide 108).

```
batch_norm_layer = None
for module in model.modules():
    if isinstance(module, torch.nn.BatchNorm2d):
        batch_norm_layer = module
        print(batch_norm_layer.bias.data)
        print(batch_norm_layer.weight.data)
```

```
tensor([-0.6594, -0.2657, -0.5547, -0.4734, -0.1127, -0.4061, -0.1436, -0.2137,
        -0.5908, -0.4748, -0.3164, -0.3579, -0.5057, -0.4867, -0.7317, -1.0572])
tensor([0.5830, 0.1399, 0.8804, 0.9056, 0.7414, 0.9850, 0.5473, 0.4380, 0.5924,
        0.6812, 0.5822, 0.5522, 1.0621, 0.8961, 0.7337, 0.2937])
tensor([-0.1509,  0.0959, -0.3734, -0.2218, -0.3429, -0.1273, -0.1207, -0.4595,
        -0.0101, -0.8384, -0.0601, -0.3529, -0.3811, -0.0982,  0.0717, -0.3933])
tensor([0.5406, 0.5034, 0.3642, 0.4720, 0.5952, 0.6006, 0.6269, 0.6198, 0.4370,
        0.7629, 0.4952, 0.5131, 0.5299, 0.6316, 0.3576, 0.1313])
```

d) [8pts] We want to see the impact of batch size on the performance of the model. For the same learning rate and number of batches, try the following values of batch size [32, 64, 256]. For each

6

batch size, please plot the accuracy for both test and training data (You should thus provide three (3) plots). The variables have already been provided. You should just run and plot them. As discussed, add your name and student ID in the title.

e) [2pts] How do you compare the plot with batch size = 1 (attached) vs batch size = 256. How would you explain the performance issue of with batch size =1.
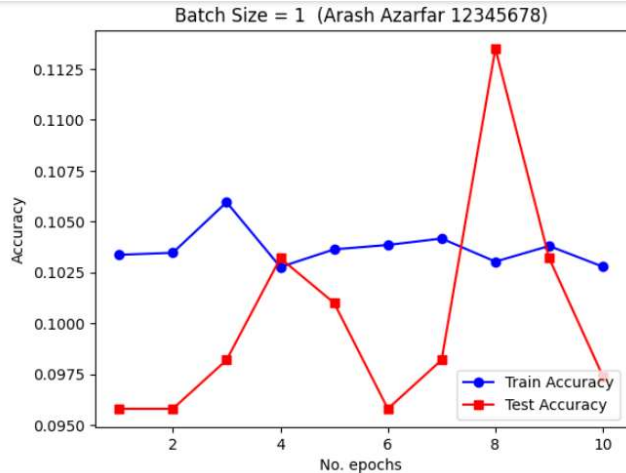
f) [3pts] In the plot with batch size = 256, how do you interpret the difference between the training and test accuracy? How may you mitigate the issue? How may you change the structure of CNN to mitigate it?

Question 6 [10pts] News Dataset

We will use the dataset of news category of HuffPost available here on Kaggle:

https://www.kaggle.com/datasets/rmisra/news-category-dataset

Please reference the dataset in your submissions. If you use Python or your notebook locally, please extract and put the file *News_Category_Dataset_v3.json* in the same folder as the python file or notebook. If you use Google Colab, you should first upload the file to Google Colab.

In the provided notebook, some steps are already written

for you. You complete the code and answer the following questions.
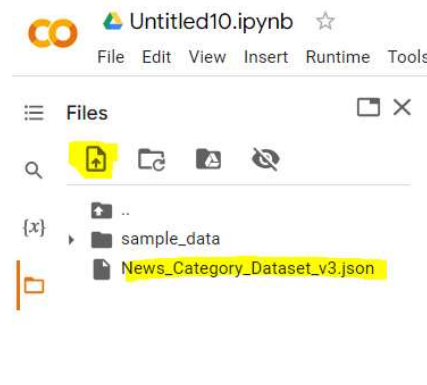
a) [3p] Create a bag of words using `CountVectorizer` Sickit-Learn from the headline column (as we did in the labs). What is the size of the corpus (no. of words) if you do not remove stop words? What is the size if we remove English stop words? What are the first three most frequent words in this corpus?

b) [5pt] Remove the stop words in the `CountVectorizer`. As we did in the lab, calculate the tfidf weights of every sentence in the dataset. Then, provide your own sentence, find the tfidf weight of your sentence, and then find which sentence in the dataset is the most similar to yours, based on Cosine Similarity. Take a snapshot of your own sentence and the one in the dataset and the Cosine Similarity.

Example:

```
test_sentence
'I do not understand this assignment'
```

```
[14] import numpy as np
     idx = np.argmax(similarity_scores)
     idx

     57285
```

```
[ ] df.iloc[idx,:]['short_description']
```
```
'When presenting the outcome of your assignment, it is important to choose the right elements of the way y
ou handled the assignment. If you don't, you face the risk of being taken for granted, not getting enough
credit for the assignment OR just simply taken advantage of your work.'
```

c) [2pts] Use Spacy to analyze the name entities (NE) in the new headline (the one which was the most similar to yours) (the same as we did in the lab]. Provide a snapshot of the answer.

Question 7) [5p] Compare CONV architectures.

For a network with an input layer of size (64x64x3) (i.e., RGB color images of size 64x64), compare the following two architectures:

- Three consecutive convolution layers each with 3x3 filters, after the input layer
- One convolution layer with 7x7 filters, after the input layer

Compare output volume size, feature detection capacity, and number of parameters.

Which one do you recommend and why?

Hint: (Due to parameter sharing in convolutional networks and the fact that the same image size is input to both architectures, the size of the images is not needed to be considered)