CMPE 353 AI: Final Examination
4 FEB 2022 1 hour

Soltion

Name & Surname:

ID:

Signature:

PART- A

! Please mark your answer in the table

(each question in PART-A is 5 points)

1	2	3	4	5	6	7	8	9	10
A	D	C	A	C	D	E	B	E	B

Q1) Which one would be the worst (not proper) design of CNN to solve any Computer Vision problem *(**I**: Input, **F**: Feed Forward (Fully Connected) **C**: Convolution Layer, **P**: Pooling)* ?

A) I-> F -> P-> F->C
B. I-> C -> P -> C-> P->F
C. I-> C -> P-> C->F
D. I-> C -> C-> P-> C->F
E. I-> C -> P-> P -> C->F

Q2) What would be the typical length range (vector size) of word embeddings?

A) 1-4 B) 4-8 C)8-32 D) 32-512 E) 4096+

Q3) Which one ignores the sequential order of words?

A. CNN
B. RNN
C. Bag Of Words (BoW)
D. LSTM
E. All of them

Q4) How would you rank the approaches in ascending (from worst to best) order by their performance in general?

A. BoW < RNN =CNN < Transformers
B. BoW < CNN < Transformers < RNN
C. RNN =CNN < Transformers < BoW
D. RNN < BoW<CNN < Transformers
E. Transformers<RNN<CNN < BoW

Q5) How do we produce word embeddings?

A. We produce the embeddings of the words by their dictionary definitions
B. We count how many times a word co-occur other words
C. We learn the embeddings of the words by training a network using their neighbor's words
D. Randomly produced vector.
E. We manually define them in order to maximize the log-likelihood

Q6) Which one is **incorrect** about training a deep learning model?

A. We pass dataset many times (epochs) through neural nets
B. Training performance tends to be better than test performance
C. Training too much leads to better training performance
D. Training too much leads to better test/validation performance
E. We start training with random weights of neural nets.

Q7) What do we transfer in ***"transfer learning"***?

A. Architecture only (number of layers)
B. Hyperparameters (learning rate, model type, optimizer, etc.)
C. Training data
D. Model Predictions
E. Model weights

Q8) We want to retrieve the university name (Bilgi University, Stanford University etc.) from a list of documents. Which model would fit the problem most?

A. Text Classification
B. Named-Entity Recognition
C. Sentiment Analysis
D. Text Clustering
E. Table Question Answering

Q9) What is zero-shot learning?

A. We initialize the weight of a neural network with all zero
B. We do not use any weight (zero- weights)
C. We do not use any hyper-parameter
D. It is a situation when the model poorly fits and gets (0) zero success
E. Just use the model without training on any training example

Q10) What is eXplainable AI (XAI)?

A. We explain which type of model that we used and the reason to the customers
B. Understand why a model makes such a decision when predicting
C. To visualize the weights of a big model at scale
D. Explain the mechanism of how a model learns from data
E. Forcing the model towards making better predictions

PART -B

Q11) (10 points)

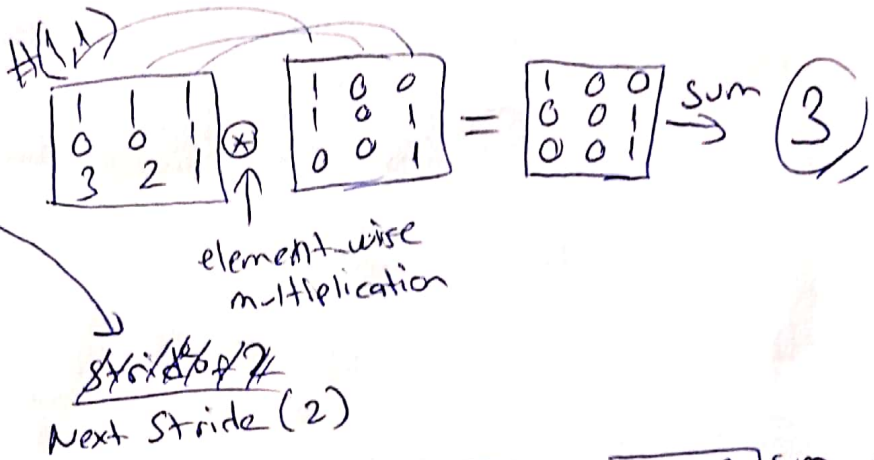
(With ***valid padding*** and the ***stride of 2***.)What would be the output when applying the filter to the input matrix below?

Input Matrix

1	1	1	2	2	1
0	0	1	1	2	1
3	2	1	1	1	0
1	1	0	0	1	1
1	1	1	0	1	1
1	0	1	0	2	3

Filter

1	0	0
1	0	1
0	0	1



The Result Matrix

3	5
5	3

Q12) What is data augmentation?
Answer (one sentence): increase data amount by a model

Q13) Give two examples to augment the data for computer vision problems!

Example 1: Mirror an image

Example 2: Zoom-In and Crop

Q14) Suppose we trained a model and we got the following results in 10 epochs. The first row represents the first epoch performance. (loss: Training Loss, acc: Training Accuracy, Val means *validation*.) At each epoch, we save the checkpoints.

```
- loss: 0.4838 - acc: 0.7481 - val_loss: 0.3685 - val_acc: 0.8272
- loss: 0.3211 - acc: 0.8519 - val_loss: 0.3336 - val_acc: 0.8435
- loss: 0.2790 - acc: 0.8723 - val_loss: 0.3164 - val_acc: 0.8561
- loss: 0.2497 - acc: 0.8849 - val_loss: 0.3018 - val_acc: 0.8584
- loss: 0.2269 - acc: 0.8959 - val_loss: 0.2993 - val_acc: 0.8656
- loss: 0.2099 - acc: 0.9047 - val_loss: 0.2930 - val_acc: 0.8687
- loss: 0.1955 - acc: 0.9090 - val_loss: 0.3031 - val_acc: 0.8692
- loss: 0.1835 - acc: 0.9148 - val_loss: 0.3053 - val_acc: 0.8711
- loss: 0.1744 - acc: 0.9179 - val_loss: 0.3038 - val_acc: 0.8694
- loss: 0.1655 - acc: 0.9223 - val_loss: 0.3120 - val_acc: 0.8633
```

We select the row with the min val loss

According to the result, which checkpoint do you use for production?

6th Checkpoint

Why (one sentence):

Due to Min val loss score

Q15) How do you process the following sentence in an RNN-like model? Please answer it by drawing RNN architecture and placing sentence properly.

"All you need is to work"