CS381/780 Machine Learning Final Exam 5/23/2022

Instruction: For multiple choice questions, clearly circle one of the choice; for all other questions, write your answer right below the questions. All questions carry the same weights.

Name: Jonathan Yulan-Cogollo

Question 1: Jack, Joe and John work as a data engineer, data analyst and data scientist respectively for a hedge fund respectively. Mary is a Portfolio Manager. Who she will contact when she wants to (1) get a list buy recommendation from internal research (2) summarize who has been the past 3 months winners and (3) figure out why there is a an inconsistence between the earning reports between her records versus what are reported from Bloomberg?

   A. John, Joe and Jack respectively
   B. Joe, Jack and John respectively
   C. Jack, Joe and John respectively
   D. John can handle all tasks
   E. Jack can handle all tasks

Answer: A

Question 2: Chris is a data scientist working for home insurance department of Geico. The goal of his model to predict when Geico will need to pay out a big sum of money for flooding so it can charge the right amount of premium. Which of the following statements is/are true regarding his models?

```
1. His model commits a Type-1 error when it predicts Geico needs to make a big
payout in 3 months.
2. His model commits a Type-2 error predict Geico does not need to pay but at the
end flooding did happen.
3. He prefers his model to have a high precision than a high recall.
```

   A. Only 1   B
   B. Only 2
   C. 1 and 2
   D. 2 and 3
   E. All of the above

Question 3: Which of the following statements related to Linear and Logistic regression is/are true?

> 1. The mean square error function can be used as the cost function for
> Logistic regression.
> 2. Regression coefficients are found by minimzing the corresponding cost
> function.
> 3. The range of the sigmod function is between -1 and 1.

A. Only 1
B. Ony 2
C. 1 and 2
D. 1 and 3
 E. All are above

Answer: B

Question 4: Which of the following statements related to Linear and Logistic regression are true?

> 1. To fit a model, we try to find the regression coefficients by minimizing a
> certain cost function.
> 2. Principle of maximum likelihood can be used to find the best fitted Linear
> regression model.
> 3. F1-score can be used to measure the goodness of a fit for both Linear and
> Logistitc regression model.

A. Only 1
B. Ony 2
C. 1 and 2
D. 2 and 3
 E. None of the above

Answer: A

Question 5: Which of the following are true regarding decision trees algorithm?

> 1. Random forest is based on divide-and-conquer technqiue.
> 2. Decision trees algorithm is based on ensemble methods where an emsemble of
> all its features are considered.
> 3. We want to pick an attribute that decreases entropy the most.

A. Only 1
B. Only 2
C. Only 3
D. 1 and 3
 E. 2 and 3

Answer: D

Question 6: Which of the following is/are true regarding Machine Learning?

> 1. Machine Learning is the study of how machine can simulate human behavior in solving problems.
> 2. Deep learning will automatically discover the optimal combination of the inputs variables without the need of features engineering from a domain expert.
> 3. Shallow Machine Learning hypothesis space is more extensive than what is allowed in Deep Learning.

A. Only 1
B. Only 2
C. 1 and 2
D. 2 and 3
E. All of the above

Answer: B

Question 7: Answer the following questions based on the following results from a classification model

n = 100	Actual Yes	Actual No
Predicted Yes	35	6
Predicted No	4	55

A. Calculate the Precision: Your answer: TP/(TP + FP) = 35/ (35 + 6) = 0.854

B. Calculate the Recall: Your answer: TP/(TP + FN) = 35/ (35 + 4) = 0.897

C. Calculate the True Positive Rate: Your answer: TPR = RECALL = 0.897

D. Calculate the False Positive Rate: Your answer: FPR = FP/ (FP + TN) = 6 / (6 + 55) = 0.098

Question 8: Which of the following regarding Support Vector Machine (SVM) are true?

> 1. The decision boundary is a hyperplane but can be a straight line in 2-dimensional or 3-dimensional feature space.
> 2. SVM finds its optimal model parameters by maximizing the margin between the decision boundary and the observed data.
> 3. The difference between Soft margin and Hard margin is that soft margin allows some mis-classification on some data points.

A. Only 2

B. Only 3

C. 1 and 2

D. 2 and 3

E. All of the above

Answer:

B

Question 9: Which of the following is/are true regarding Deep Learning Libraries?

```
1. TensorFlow is a Python-based platform, but has been exported to C++ and
JavaScript environment.
2. Keras was originally designed specifically as a top-level API for TensorFlow to
hide complexiity of building a network from scrtach.
3. All PyTorch, TensorFlow and CNTK are open-source libraries for Deep Learning.
```

A. Only 1

B. Only 3

C. 1 and 2

D. 1 and 3

E. All of the above

Answer:

D

Question 10: Which of the following is/are true regarding activation functions?

```
1. Purpose of the activation functions is to introduce linear effects
2. Sigmoid function is more preferable than other non-continous function as it is
differentiable everywhere.
3. ReLu can be used as an activation function, but not Leaky ReLu as it leaks and
loses information
```

A. Only 1

B. Only 3

C. 1 and 2

D. 1 and 3

E. None of the above

Answer:

E

Question 11: Which of the following is/are true regarding technqiues in avoiding overfitting?

```
1. Data Augmentation is a technique to draw additional samples from the input
training set to help reduce overfitting.
2. Learning curves allows you to investigate whether you have an underfit case.
3. The goal of using Dropout to help avoid overfiting and it is guaranteed to
work.
```

A. Only 1
B. Only 2
C. Only 3
D. 1 and 2
E. All of the above

Answer: B

Question 12: Which of the following is/are true regarding Ensemble Learning methods applied to classificatoin problem?

1. Training process based on Bagging can be run in parallel but not processes
based on pasting because pasting involve sampling in pasting is done without
replacement.
2. Performance of Hard Voting will often outperform Soft Voting as it is based on
the majority.
3. AdaBoost focus on predicting the residual error at each iterations.

A. Only 1
B. Only 2
C. 1 and 2
D. 2 and 3
E. None of the above

Answer: E

Question 13: Answer the following questions regarding Vector, Matrix and Tensor.

(A) What is the rank of the tensors required for storing images data for Deep Learning?
Answer: tensor of rank-4

(B) What kind of data the following diagram represents? Describe a corresponding example dataset.

This is an example of image data (4D), it has (#_samples, height_of_images, width_of_images, channels);
A batch of 9RBG images of size 800 x 800 could thus be stored in a tensor of shape (9,800,800,3) where the values
are represented by the labels I have listed above. The 9 is what I can count, not sure if there are more, and this is
colored thus there is RGB code represented by the #3.
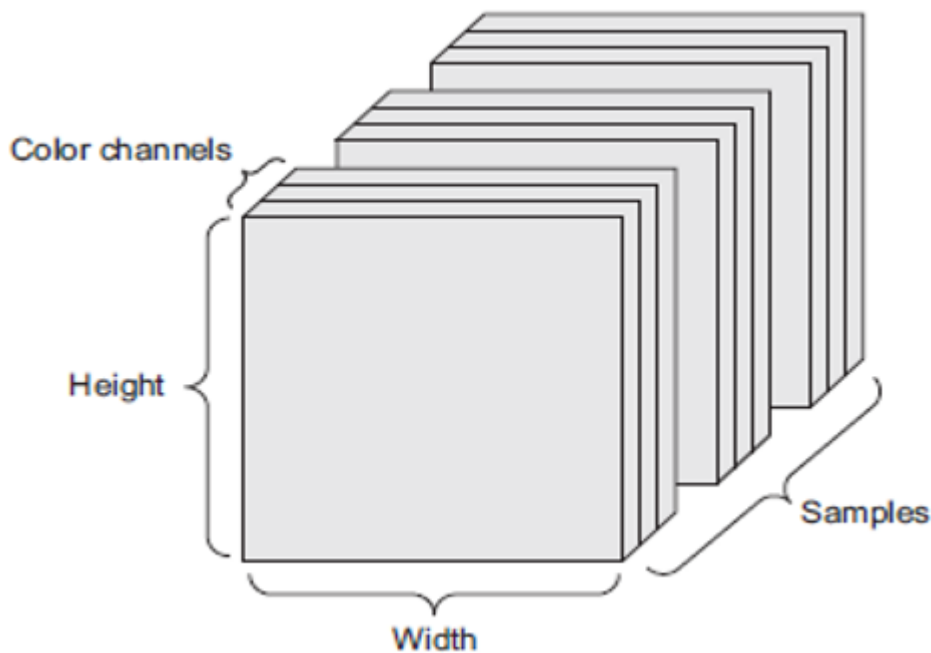
Answer:

(C) The following 2 vectors, v_1, v_2, are samples drawn from a dataset.

$$v_1 = (1, 3, -2, 1)$$

$$v_2 = (-1, 1, 1, 3)$$

What is the dimension of these vectors? What is the rank of these vectors as tensors? Compute the dot product of v_1 and v_2

Answer: 4 Dimensional Vector   (1)(-1) + (3)(1) + (-2)(1) + (1)(3) = 3
          1 Rank Vector

Question 14: Which of the following is/are true regarding Recurrent Neural Network (RNN) and Convolution Neural Network (CNN)?

```
1. The dependence between the future inputs at time t on the inputs node at time
zero increase with time t because of accumulation effects.
2. One of the main difference between a RNN with CNN is that RNN keeps track of an
internal state but CNN does not.
3. Both LSTM and GRU has a forget gate for removing non-importance information in
their memory.
```

A. Only 1
B. Only 2
C. 1 and 2

D. 2 and 3
E. All of the above

Answer: B

Consider the problem of classifying images into dog, cat or duck, the following table summarizes the probabilities of each class from 4 different models on a randomly selected image from a testing dataset with equal number of dog, cat and duck images.

Class	Model A	Model B	Model C	Model D
Dog	0.3	0.05	0.45	0.33
Cat	0.3	0.75	0.4	0.33
Duck	0.4	0.2	0.15	0.34

?????????

Question 15: On given the above table, rank the cross entropy of the models in ascending order. Choose from one of the following choice

```
1. A < B < C < D          SUMMATION P(X) LOG2(X)
2. A < D < C < B
3. D < A < B < C
4. D < A < C < B
5. B < C < A < D
6. None of the above.
```

Answer: D

Question 16: On given the above table, calculate what will be the predicted class (Dog, Cat or Duck) from an voting classifier that consists of all the 4 models using (A) hard voting or (B) soft voting. Explain your reasoning.

Answers: (A)

Hard Voting is done when the aggregation is done through each predictor's output class.
CAT is what I choose for this because .3+.75 + .4 + .33 = 1.78 / 4 => .445 which was the largest value I received for the probabilities for the predicted classes.
(B)

Duck. Soft Voting is done when the aggregation is done utilizing the output probabilites.

Question 17: What are the size of the output feature map at the end of each layers if size of the input images is 68 x 68?

layer 1: Conv2D(filters=32, kernel_size=3, activiation="relu")
layer 2: MaxPooling2D(pool_size=2)
layer 3: Conv2D(filters=32, kernel_size=7, activiation="relu")
layer 4: MaxPooling2D(pool_size=3)

Answer:

layer 1:   66
layer 2:   33
layer 3: 27
layer 4:   13

Question 18: Given the following image pixels and filter, write down convolved feature map

Image pixels

1	0	0	1	0
0	0	1	1	0
1	1	1	0	0
1	1	0	0	0
0	0	1	0	1

Filter

0	1	1
1	0	1
0	1	0

2 | 3 | 2
————
4 | 3 | 2
————
3 | 3 | 0

Answer:

Question 19: Convolution of functions f and g is defined by

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t - \tau)g(\tau)d\tau$$

It is what the standard convolution operation in Convolution Neural Network uses with f as the image and g being the function represented by the filter.

(A) Now if we define a "Enhanced Convolution" operator by

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t - \tau)(g(\tau) + h(\tau))d\tau$$

where h is a so-called "enhancer" filter.

Compute the corresponding "Enhanced convolved" feature map result by using the same image and filter from the previous question and the following enhancer filter h

Enhancer Filter

1	0	0	(0)+(1) | (1) + (0) | (1) + (0)	1| 1 | 1
0	1	0	(1)+(0) | (0) + (1) | (1) + (0)	1 | 1 | 1
0	0	1	(0)+(0) | (1) + (0) | (0) + (1)	0 | 1 | 1

(B) Now if we define a "Extended Convolution" operator by

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t - \tau)g(\tau) + h(\tau)d\tau$$

where h is a so-called "extender" filter.

Compute the corresponding "Entended convolved" feature map result by using the same image and filter from the previous question and the following extender filter h.

Extender Filter

0	0	0
0	1	0
0	0	0

Answer: (A)

4 | 4 | 3

5 | 4 | 3

6 | 4 | 2

(B)

2 | 3 | 2

4 | 4 | 2

3 | 3 | 0

Question 20: In your own words, describe the vanishing gradient problem and one example where it may appear.

Answer:

Normally utilizing the sigmoid activation function you get very, very small gradients at the ends of your functions. When you combine this with a very deep model, with many layers, which is commonly found in a RNN model or one that keeps a resemblance of memory the sensitivity of the model's parameters cannot be reliably calculated due to the gradient becoming too small. When you get these very small numbers, you create an almost zero change when going from iteration to iteration within your model and thus the model weights in your epochs begin to stall, and you become stuck, causing your model to stall. Normally this is found in very deep layers of neural networks maintaining several dimensions that cause the parameters to become very small and stall.