CS583A: Final Exam (Sample Questions)

Name:

Policy: Books and printed materials are allowed. Do not use electronic divice, including phone, laptop, and tablet.

Hint: (i) $\frac{\partial e^a}{\partial a} = e^a$, (ii) $\frac{\partial \log_e(a)}{\partial a} = \frac{1}{a}$, (iii) $\frac{\partial \frac{1}{a}}{\partial a} = -\frac{1}{a^2}$, and (iv) $\frac{\partial \cos(a)}{\partial a} = -\sin(a)$.

Q1 (2%). Let $\mathbf{A}$ be a real matrix and $\mathbf{b}$ be a real vector. To compute the multiplication $\mathbf{Ab}$ efficiently, which of the following libraries is the the best choice?

A. Level 1 BLAS.

B. Level 2 BLAS.

C. Level 3 BLAS.

D. LAPACK.

Q2 (2%). The input shape is 18×18, the pool size is 3×3, and the pooling has no overlap (equivalently, the stride is 3×3). Then what is the output shape of the pooling?

A. 3×3.

B. 6×6.

C. 15×15.

D. 16×16.

E. 18×18.

Q3 (2%). Let $\mathbf{a} = [1, 2, 3, 4, 5]^T$ be a vector. Let $\mathbf{b} = \tanh(\mathbf{a})$, where tanh is the hyperbolic tangent function. Then $\|\mathbf{b}\|_2^2 < 1$.

A. The statement is true.

B. The statement is false.

Q4 (2%). We want to train a deep convolutional neural network using the CIFAR10 dataset. Which of the following tricks cannot be applied to improve the test accuracy?

A. Using dropout regularization.

B. Using data augmentation.

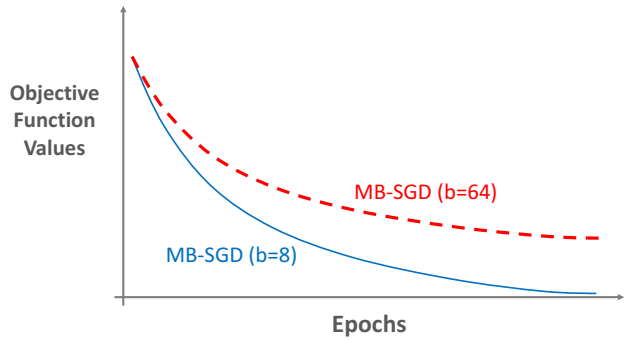
C. Using ensemble method.

D. Using multi-task learning.

Q5 (2%). We use mini-batch SGD with momentum to solve a softmax classifier model. We use different settings of batch size, b, and plot the **objective function value** against **epochs** in the figure below. All the other hyper-parameters in the algorithms have been fine-tuned. The figure indicates that MB-SGD with $b = 8$ is very likely better than $b = 64$ in this case.

A. The statement is true.

B. The statement is false.



Q6 (2%). Suppose we seek to predict increment (either positive or negative) of a stock's price based on its prices in the past. Which of the followings is the best choice for the activation function of the output layer?

A. No activation function (equivalently, the identity function).

B. Rectified linear unit (ReLU).

C. Logistic function.

D. Softmax function.

D. Tanh function.

Q7 (2%). We use a deep learning model to predict whether a customer's review is positive or negative. We tokenize the reviews in the word-level and thus need an embedding layer to convert words to vectors. Upon the embedding layer, we build RNN layers and other types of layers. We find serious overfitting because the dataset has merely 10,000 samples. Which of the followings is the best improvement?

A. Pretraining the embedding layer using a big dataset.

B. Using Bi-LSTM to replace SimpleRNN or LSTM.

C. Combing self-attention with RNN.

D. Using stacked-LSTM to increase the model capacity.

Q8 (2%). The following implements a convolutional neural network. The code has a major problem. How can we make correction?

A. Remove the second convolutional layer (Lines 10 to 12), because it is not useful.

B. Put the batch normalization layers (Lines 8 and 11) after the activations (Lines 9 and 12).

C. Put the batch normalization layers (Lines 8 and 11) before the convolutions (Lines 7 and 10).

D. Put the dropout layer (Line 18) before the first dense layer (Line 16).

E. Put the dropout layer (Line 18) before the second dense layer (Line 17).

```
1  from keras import models
2  from keras.layers import Input, Conv2D, MaxPooling2D, Activation
3  from keras.layers import BatchNormalization, Flatten, Dense, Dropout
4
5  input_img = Input((128, 128, 3))
6
7  x = Conv2D(32, (3, 3))(input_img)
8  x = BatchNormalization()(x)
9  x = Activation('relu')(x)
10 x = Conv2D(32, (3, 3))(x)
11 x = BatchNormalization()(x)
12 x = Activation('relu')(x)
13 x = MaxPooling2D((2, 2))(x)
14
15 x = Flatten()(x)
16 x = Dense(128, activation='relu')(x)
17 x = Dense(256, activation='relu')(x)
18 x = Dropout(0.5)(x)
19 x = Dense(10, activation='softmax')(x)
20
21 model = models.Model(input_img, x)
```

Q9 (2%). The following implements a convolutional neural network. The code is incorrect. How can we make correction?

A. The activation in Line 7 should be removed.

B. The activation layer (Line 9) should be removed.

C. The batch normalization layer (Line 8) should be place after the pooling layer (Line 10).

D. The dropout layer (Line 12) should be placed after Line 13 and before Line 14.

```
1  from keras import models
2  from keras.layers import Input, Conv2D, BatchNormalization, Activation
3  from keras.layers import MaxPooling2D, Flatten, Dropout, Dense
4
5  input_img = Input((28, 28, 3))
6
7  x = Conv2D(32, (3, 3), activation='relu')(input_img)
8  x = BatchNormalization()(x)
9  x = Activation('relu')(x)
10 x = MaxPooling2D((2, 2))(x)
11 x = Flatten()(x)
12 x = Dropout(0.5)(x)
13 x = Dense(1000, activation='relu')(x)
14 x = Dense(10, activation='relu')(x)
15
16 model = models.Model(input_img, x)
```

Q10 (2%). (Fill the blank.) The training set contains $1,000$ samples. We train a neural network using mini-batch SGD. One epoch amounts to 50 iterations. What is the batch size?
The batch size is _____.

Q11 (2%). (Fill the blank.) What is the output of the following Python program?
Answer: _____

```
1 import numpy
2 a = numpy.random.rand(3, 5) # generate a random matrix
3 b = numpy.random.rand(5, 10) # generate a random matrix
4 c = numpy.dot(a, b)
5 print(c.shape[0])
```

Q12 (12%). (Fill the blanks.) The following code builds a neural network for sentiment analysis.

- Line 10 is an embedding layer. What is the output shape of this layer?
  Answer: _____

- What is (roughly) the number of parameters in the embedding layer?
  Answer: _____
- Line 11 is a SimpleRNN layer. What is the output shape of this layer?
  Answer: _____
- What is (roughly) the number of parameters in the SimpleRNN layer?
  Answer: _____
- Line 12 is a dense layer. What is the output shape of this layer?
  Answer: _____
- What is (roughly) the number of parameters in the dense layer?
  Answer: _____

```
1  from keras.models import Sequential
2  from keras.layers import SimpleRNN, Embedding, Dense
3
4  voc_size = 10000
5  shape_x = 20
6  seq_length = 100
7  shape_h = 50
8
9  model = Sequential()
10 model.add(Embedding(voc_size, shape_x, input_length=seq_length))
11 model.add(SimpleRNN(shape_h, return_sequences=False))
12 model.add(Dense(1, activation='sigmoid'))
```

Q13 (12%). The following code builds a convolutional neural network.

- Line 5 is a convolutional layer with 1×1 convolutions. What is the output shape of this layer?
  Answer: _____
- What is (roughly) the number of parameters in the convolutional layer (Line 5)?
  Answer: _____
- Line 8 is a convolutional layer with 3×3 convolutions. What is the output shape of this layer?
  Answer: _____
- What is (roughly) the number of parameters in the convolutional layer (Line 8)?
  Answer: _____
- Line 13 is a pooling layer. What is the output shape of this layer?
  Answer: _____
- What is (roughly) the number of parameters in the pooling layer (Line 13)?
  Answer: _____

```
from keras.layers import Input, Conv2D, MaxPooling2D, concatenate

x_input = Input(shape=(100, 100, 40))

tower1 = Conv2D(10, (1,1), padding='same', activation='relu')(x_input)

tower2 = Conv2D(10, (1,1), padding='same', activation='relu')(x_input)
tower2 = Conv2D(10, (3,3), padding='same', activation='relu')(tower2)

tower3 = Conv2D(10, (1,1), padding='same', activation='relu')(x_input)
tower3 = Conv2D(10, (5,5), padding='same', activation='relu')(tower3)

tower4 = MaxPooling2D((3,3), strides=(1,1), padding='same')(x_input)
tower4 = Conv2D(10, (1,1), padding='same', activation='relu')(tower4)

x_output = concatenate([tower1, tower2, tower3, tower4], axis = 3)
```

Q14 (6%). A question based on the textbook.

Q15 (10%). A matrix calculus question analogous to those in the quiz.