

JK Lakshmipat University Jaipur
INSTITUTE OF ENGINEERING AND TECHNOLOGY

End Term Examination, 27 May 2022

B.Tech., Semester VI, 2021-22

CS1218: Deep Learning

Time: 2:30 Hours

Roll number

Max. Marks: 40+15

Read the following instructions carefully:

- There are 5 questions. The paper is printed on both sides.
- If in doubt, state your assumption(s) and continue.
- **No marks** unless justification and/or derivation accompanies the answer.

1. (a) Why is Mean Squared Error used as a cost function for linear regression and cross-entropy for logistic regression? Write their expressions.
- (b) Assuming four input features are (x_1, x_2, x_3, x_4) , draw a linear regression network and a logistic regression network. Neatly label all the elements of both the networks along with their input and output.
- (c) Prove that the logistic regression is limited to linear classifiers only.

(2+2)+(2+2)+2 = 10 Marks, [CO4]

2. Assume a huge text corpus whose vocabulary size is 10,000 words.

- (a) How will you represent each word using one hot vectors?
- (b) How will the representation change if the corpus has 2,00,000 words in the vocabulary?
- (c) State two prominent disadvantages of one hot vector representation.
- (d) What are word embeddings?
- (e) How do word embeddings solve the problems mentioned by you in part (c)?

2+2+2+2+2 = 10 Marks, [CO1, CO2]

3. (a) In a convolutional neural network (CNN), if the filter size is $f \times f$, the size of the image is $n \times n$, padding used is p and a stride of s is also applied. Derive a formula for the dimensions of the output.
- (b) Let the input volume to a convolutional layer have dimensions $6 \times 6 \times 3$. You want to have convolution with a filter that has size $3 \times 3 \times 2$. Suggest a way to be able to do so. Note that the number of channels are not matching and we have learnt it in the classes that cheap tricks work! Hence, provide a cheap trick of your own! Also find the dimensions of the output if no padding is used and stride is 1.

5+(3+2) = 10 Marks, [CO2]

4. Why is deep learning taking off right now? Draw the figure that we had seen in the class on *Performance vs Amount of data*. In the context of the figure, elaborate your answers based on the following:

- (a) Computational resources
- (b) Existence of data
- (c) Existence of the Internet
- (d) Advances in the field itself

2+2+2+2+2 = 10 Marks, [CO3]

5. Analyze the following piece of code.

```
import tensorflow as tf
import numpy as np
import matplotlib.pyplot as plt

mnist = tf.keras.datasets.mnist
(x_train, y_train), (x_test, y_test) = mnist.load_data()

x_train=x_train.reshape(60000, 28, 28, 1)
x_train=x_train / 255.0
x_test = x_test.reshape(10000, 28, 28, 1)
x_test=x_test/255.0

model = tf.keras.models.Sequential([
    tf.keras.layers.Conv2D(32, (3,3), activation='relu', input_shape=(28, 28, 1)),
    tf.keras.layers.MaxPooling2D(2, 2),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(10, activation='softmax')
])

model.compile(optimizer='adam', loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

model.fit(x_train,y_train,epochs=1)

model.evaluate(x_test, y_test)
```

- Draw the neural network corresponding to the model in this code.
- Assuming there are 5 layers in the above model (one corresponding to each `tf.keras.layers` statement), what is the number of parameters that can be learnt at each layer?
- How does the answer to 5(b) change if `MaxPooling2D(4,4)` is used instead of `MaxPooling2D(2,2)`?
- If the model in the above code is changed to the following. Then find the number of learnable parameters in the model.

```
model = tf.keras.models.Sequential([
    tf.keras.layers.Flatten(input_shape=(28, 28)),
    tf.keras.layers.Dense(256, activation=tf.nn.relu),
    tf.keras.layers.Dense(10, activation=tf.nn.softmax)
])
```

4+5*1+2+4 = 15 Marks, [CO4, CO5]