CS550: Machine Learning Tierce Examination 2

Name:

Roll Number:

Instructions

a. Read all questions carefully before you start to answer them. In case of any doubt, you can write your assumptions before answering the question.
b. This is a CLOSED BOOK exam.
c. You can use a calculator.
d. Use of Phone, INTERNET or Google Searches etc. is NOT ALLOWED.
e. Discussion with any other person is NOT ALLOWED. Any unfair means used will result in D or F grade and will be reported to the competent authority.
f. You will get 90 minutes to solve all the questions. We won't be giving extra time. Please submit whatever you have done at the end of the allotted time. Anyone found writing after completion time will get -20 marks.
g. PLEASE ALWAYS PROVIDE A JUSTIFICATION FOR YOUR ANSWERS. CORRECT SOLUTIONS WITHOUT ANY JUSTIFICATION WILL NOT GET FULL CREDITS. THEY MAY ALSO BE CONSIDERED AS POTENTIAL UNFAIR MEANS.
h. Make sure to write your name and roll number on the top of your sheet(s) when submitting.
i. Please write the answers in the EXAM PAPER itself. The justifications and calculations can be in the answer sheet. We will look at answer sheet only if needed.

For Instructor's use only:

Q	Topic	Max. Marks	Marks Obtained
1	Objective questions	40	
2	Product Similarity	30	
3	Segmentation	20	
Total		90	

Q1. General Concepts (2 marks each)

A. What is the core difference between the fully connected neural network and the CNN?
   a. Fully connected NN is used only for one-dimensional data while CNN is used solely for higher dimensional data.
   b. Weights are arranged in a filter that slides over the input, calculating the output piece by piece.
   c. Instead of using nonlinear activation functions, it uses linear activation functions

B. Which mathematical operation is the main component of a convolution layer?
   a. The matrix dot product
   b. The sigmoid function
   c. Singular value decomposition
   d. Taking absolute values via the ReLU function

C. What is the purpose of a pooling layer?
   a. It helps the CNN avoid overfitting by reducing the number of parameters in the model
   b. Pooling is required for a convolution layer to work properly
   c. Pooling is an extension of the convolution layers, so the weights are reinforced
   d. None of the above

D. How does dropout help a neural network avoid overfitting?
   a. It helps each neuron depend on one another, reinforcing the best results
   b. It increases the learning rate for a certain number of training steps
   c. It randomly chooses a fraction of the neurons to drop, reducing neuron co-adaptation
   d. It randomly drops certain training iterations from contributing to the overall loss

E. Why is data augmentation useful?
   a. It helps speed up training by only using important data examples
   b. It lets us enlarge image data if they are too small
   c. It lets us simulate additional data examples, which is useful when our dataset is small
   d. It increases the resolution of images so the model has an easier time training

F. What problem does batch normalization solve?
   a. Internal covariate shift
   b. Vanishing gradients
   c. Exponentially slowed down training
   d. Large model memory footprint

G. What is residual learning?
   a. Training a model in increments across many different training runs
   b. Adding small constants to each layer's weights to improve training
   c. Dropping layers from the model when their weights tend to 0
   d. The process of learning the residual function, to make identity mapping easier

H. What purpose does the checkpoint file serve?
   a. It saves all the model states from each training iteration
   b. It specifies which saved model to restore
   c. It contains the binary saved data from the model
   d. It specifies the variable names saved from the model

I. What is a difference between classification and regression?
   a. Regression cannot return negative values while classification can
   b. Classification requires the input data to be all integers
   c. Regression returns real values while classification returns class predictions
   d. Classification models are always simpler than regression models

J. What is the main cause of overfitting in deep neural networks?
   a) The model tracks even the statistical noise in the training data.
   b) The model is using too many hidden layers.
   c) The model uses too much training data.

Short Answer Type

A. [2 mark] What will be the shape of the data at the output layer after executing the following code on input data of dimension 3×2?

```
model = Sequential()
model.add(Dense(16, input_shape=(3, 2)))
model.add(Activation('relu'))
model.add(Flatten())
model.add(Dense(4))
```

   a. Your answer: …….

B. [2 marks] The output dimensions of an image of size 256×256 pixels, passing through a convolutional filter of dimensions 5×5 with padding of 2 and stride of 2 will be ……

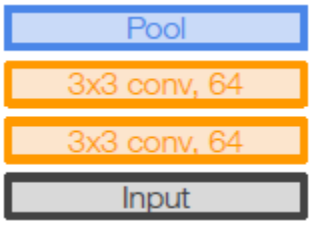
C. [2+2+4 marks] Consider vectors $\boldsymbol{u}, \boldsymbol{x} \in \boldsymbol{R^n}$, and matrix **A**$\in \boldsymbol{R^{n \times n}}$. The derivative of a scalar f w.r.t a vector **x** is a vector. Derive expressions for the following derivatives

a. $\nabla_x u^T x$

a. $\nabla_x x^T x$

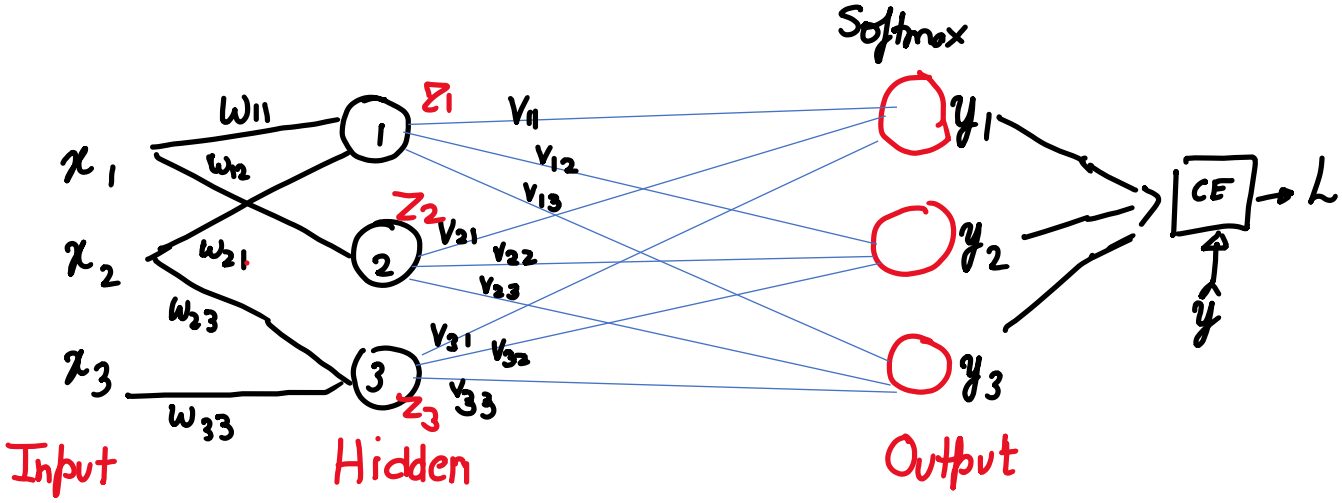
b. $\nabla_x x^T A x$

D. [8 marks] For the following block of VGG 16, compute the number of parameters and memory requirement for each layer when the input image size is 64x64x3. The 3x3 filters are used with stride 1 and padding 1, 2x2 maxpool is used with stride 2.



Q2. Product Similarity Network [30 marks]

An e-commerce company builds a feed-forward neural network to classify a product (electronics, home, baby) using its feature descriptions. The network has 1 hidden layer with swish activation function and at the output, a softmax activation function is used with cross-entropy loss.



The weights between input and Hidden layers are:

$$w_{11} = -2, w_{12} = 0.5, w_{21} = 1, w_{23} = 1, w_{33} = -2.$$

The bias is 0 and the weights that are not shown are also 0. Outputs of hidden layer are z1, z2, z3.

The hidden layer is fully connected to the output layer and the weights are given by

$$v_{11} = 1, v_{12} = -1, v_{13} = 2, v_{21} = 1, v_{22} = 2, v_{23} = -1, v_{31} = 1, v_{32} = 2, v_{33} = 3.$$

The swish activation function $S(x) = x.sigmoid(x)$

A. (5 marks) For the input of (x1, x2, x3) = (2, 4, -2), calculate the output (y1, y2, y3) and the loss if the actual $\hat{y} = (0,1,0)$

B. (5+7.5+7.5 marks) Using backpropagation, calculate the following gradients for the above input and weights.

a. $\frac{\partial L}{\partial y_1}, \frac{\partial L}{\partial y_2}$

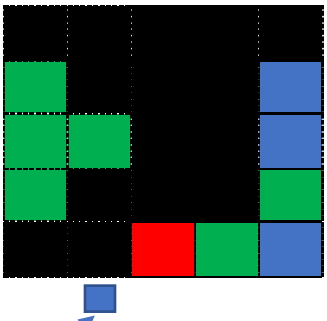
b. $\frac{\partial L}{\partial v_{11}}, \frac{\partial L}{\partial v_{21}}, \frac{\partial L}{\partial v_{31}}$

c. $\frac{\partial L}{\partial w_{11}}, \frac{\partial L}{\partial w_{12}}, \frac{\partial L}{\partial w_{23}}$

C. (5 marks) Using a learning rate of 0.1, update only the weights ($v_{11}, v_{21}, w_{11}, w_{12}$) and calculate the output and loss again.

Q3. Semantic Segmentation/ Self-driving car [30 marks]

Let us consider a toy problem for semantic segmentation of scene for a self-driving car. We are given a view from front cameras as a 5x5x3 image where road is black and obstacles (other vehicles) are colored. Your task is to distinguish between road (Class 0) and obstacle (Class 1) in this grid using a neural network.



Your car is located here.

(5 marks) Write the desired output for this particular input shown above

(15 marks) Design a Convolutional neural network that can perfectly solve the problem above. You must use a filter, perform a convolution operation and pass it through a non-linear activation function!