INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR	
	Signature of the Invigilator

Please fill up carefully the boxes provided below

EXAMINATION (Mid Semester)										SEMESTER (Spring)			
Roll Number										Section		Name	
Subject Number	C	S	6	0	0	1	0	Subject Name		Deep Learning			
Name of the Department / Center of the Student													

Instructions and Guidelines to Students Appearing in the Examination

1. Ensure that you have occupied the seat as per the examination schedule.
2. Ensure that you do not have a mobile phone or a similar gadget with you even in switched off mode. Note that loose papers, notes, books should not be in your possession, even if those are irrelevant to the paper you are writing.
3. Data book, codes or any other materials are allowed only under the instruction of the paper-setter.
4. Use of instrument box, pencil box and non-programmable calculator is allowed during the examination. However, exchange of these items is not permitted.
5. Additional sheets, graph papers and relevant tables will be provided on request.
6. Write on both sides of the answer script and do not tear off any page. Report to the invigilator if the answer script has torn page(s).
7. Show the identity card whenever asked for by the invigilator. It is your responsibility to ensure that your attendance is recorded by the invigilator.
8. You may leave the examination hall for wash room or for drinking water, but not before one hour after the commencement of the examination. Record your absence from the examination hall in the register provided. Smoking and consumption of any kind of beverages is not allowed inside the examination hall.
9. After the completion of the examination, do not leave the seat until the invigilator collects the answer script.
10. During the examination, either inside the examination hall or outside the examination hall, gathering information from any kind of sources or any such attempts, exchange or helping in exchange of information with others or any such attempts will be treated as adopting 'unfair means'. Do not adopt 'unfair means' and do not indulge in unseemly behavior as well.

Violation of any of the instructions may lead to disciplinary action of varied nature.

To be filled in by the examiner											
Question	1	2	3	4	5	6	7	8	9	10	**Total**
Marks Obtained											

Marks obtained (in words)	Signature of Examiner	Signature of Scrutineer

Indian Institute of Technology Kharagpur

Midterm 2018

Deep Learning CS60010

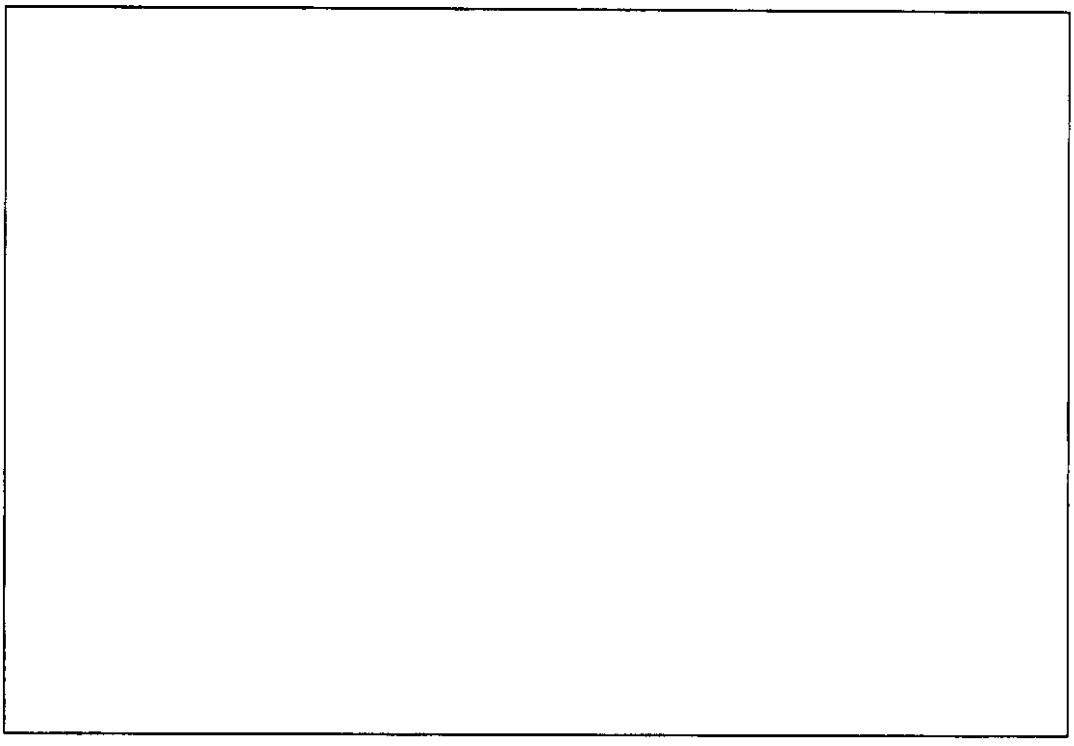
Spring 2018
Date: 22/2/2018

Midterm
Duration: 2 hours

Full Marks: 70

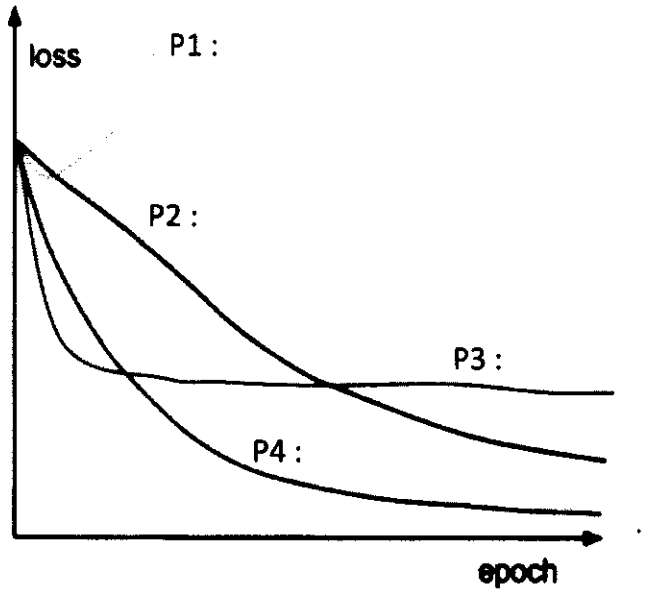
This exam contains 12 pages (including this cover page) and 4 questions.

1. (a) (4 points) Sketch the typical learning curves for the training and validation sets by plotting model error against model complexity, for a setting where overfitting occurs at some point. Assume that the training set and the validation set are of the same size. Label all the axes, and label the curves that you sketch.



(b) (4 points) The following learning rate plots show the loss of an algorithm versus time. Label each of the plots P1 to P4 with one of these labels:

A. L1: Low learning rate
B. L2: Optimal learning rate
C. L3: High learning rate
D. L4: Very high learning rate

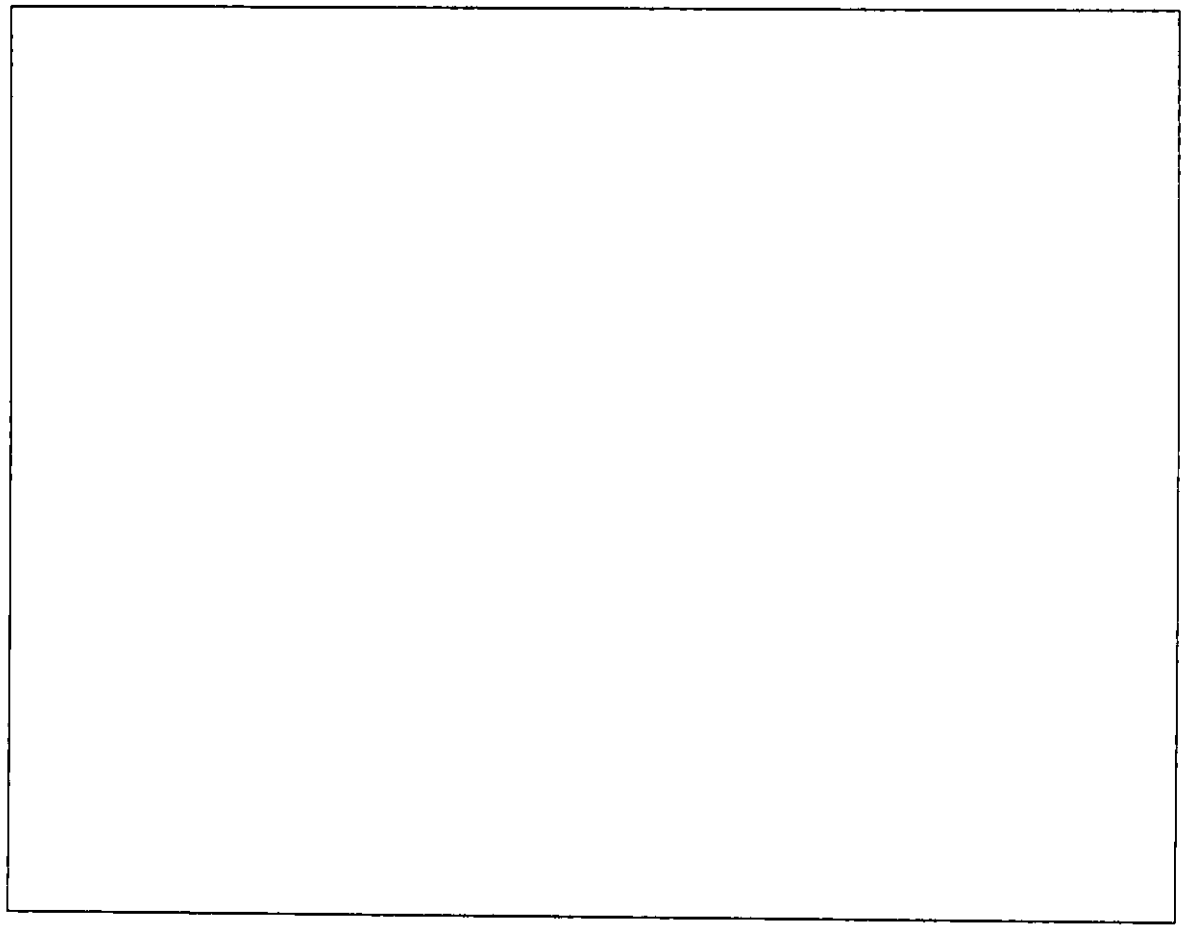


(c) (4 points) Assume a neural network with a single output using squared error. Define the symbols for all the variables required.

i. Write down the error function for online learning.

ii. Write down the error function for batch learning.

2. (a) (6 points) Briefly describe three ways of preventing overfitting in neural networks and why they work (1 line each).



(b) (6 points) Consider a neural net with three input units, 1 output unit, no hidden units and no bias terms. Suppose that the output unit uses a sigmoid activation function, i.e:, $y = \frac{1}{(1+e^{-z})}$. where z is the total input to the unit. Let y be the computed output of the neural net, let d be the desired output, and let $C = -dlogy - (1-d)log(1-y)$ be the cross entropy error. Write down the equations for a single step of weight updates by gradient descent (based on a single data sample), and derive all the necessary derivatives. Simplify your answers, and be sure to clearly identify all the variables you use.

(c) (2 points) Suppose you are training a neural net using stochastic gradient descent (SGD), and you compute the cost function on the entire training set after each update. If you ever see the training cost increase after an SGD , does that mean that the learning rate is too large? Justify your answer

(d) (6 points) Explain whether the following are true or false:

i. Adding more hidden layers will solve the vanishing gradient problem for a 2 layer neural network.

ii. Adding L2-regularization will help with vanishing gradients

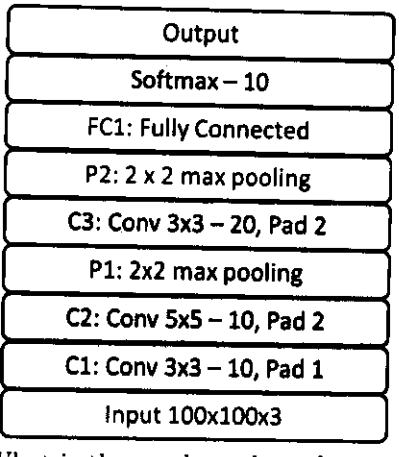
iii. Clipping the gradient (cutting off at a threshold) will solve the exploding gradients problem.

) (4 points) Describe the implementation of dropout with parameter p in a multilayer Neural Network.

i. During training

ii. During test time

oints) (a) Consider applying the Convolutional Neural Network (CNN) below to a database of images. Eacg image is of size 100×100 with 3 channels. That is, the input to the CNN has the dimension of $100 \times 100 \times 3$ and the output has 10 classes.
All convolution filters have stride 1 and all pooling functions have stride 2.



Conv $n \times n - m$, Pad k : means we apply m $n \times n$ convolution filters with padding k. All convolution filters also have bias terms.

Vhat is the total number of parameters in each of the layers?

A. C1 :

B. C2 :

C. P1 :

D. C3 :

E. P2 :

F. FC1 :

(b) (4 points) Sketch a characteristic block used in the Residual net. (The Residual net or resNet was introduced in ILSVRC 2015.) Explain the role of skip connections in this block.

4. (a) (3 points) Explain why the problem of vanishing gradients occurs in recurrent nets. Briefly explain one method of dealing with the problem.

(b) Consider a bidirectional recurrent neural network, in which there is an input x_t, an output node o_t, and two hidden nodes, $h1_t$ and $h2_t$ at each time step.

i. (3 points) Draw an equivalent, feed-forward (non-recurrent) neural network representing three time instances.

ii. (4 points) Add weights to your feed-forward network, clearly indicating which weights are equal.

(c) Consider a single LSTM Cell whose inputs are the current input x_t and h_{t-1} and C_{t-1} from the previous step; and and the outputs are h_t and C_t. Assume that the input dimension is D and the dimenion of h_t is H.

i. (2 points) Sketch a LSTM cell showing the various gates inside the cell, including *forget gate* (f_t), *input gate* (i_t), o_t, c_t and h_t.

ii. (5 points) Write down the equations of the various gates above For each of the weight parameters involved in the definitions, indicate their dimensions.

iii. (3 points) Explain why a recurrent network using LSTM cells can model long distance dependencies.

Rough Work

Rough Work

Rough Work