

Midterm exam for CPTS 434/534
Neural Networks Design And Application
Spring 2021
Friday, Mar 5, 40 mins

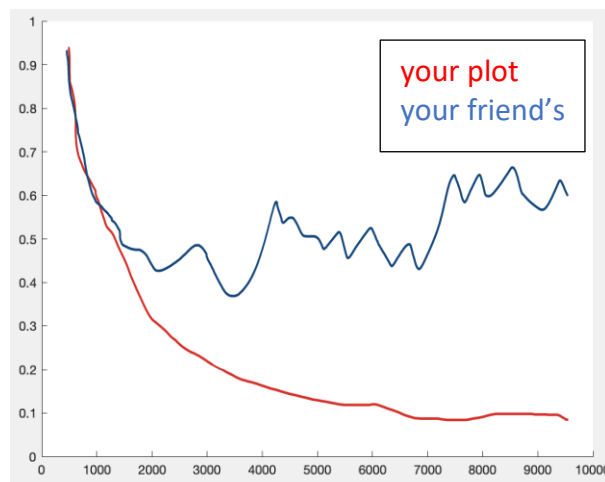
Name: Jihui.Sheng

Student number: 11539324

This is a closed book test, with up to two-page note allowed. Please answer all questions. Questions that ask you to “briefly explain” something only require short explanation that includes your main idea. For multiple choice questions, there may be more than one right answer.

Total mark: 100

1. (single choice question + explanation, 6 points) You and your friend independently trained two classifiers with identical model structure on the same training set using the stochastic gradient descent algorithm. You plot your training loss values in the following figure. What suggestion would you like to give to your friend to help the training converge? Explain why (up to 3 sentences).



- A. use a smaller learning rate (step size)
- B. continue training for a large number of iterations
- C. re-train the model from scratch with more iterations
- D. initialize the model parameters with zeros

Your answer: A, Underfitting

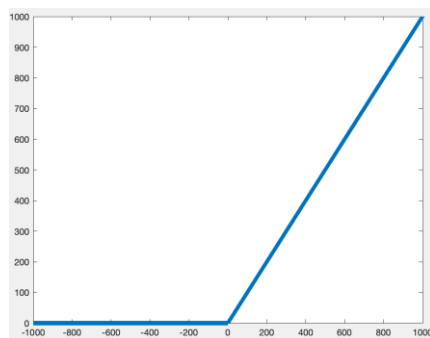
2. (multiple choice question, 5 points) Which of the following statements for logistic regression and softmax classifier are true?

- A. logistic regression is designed for multiclass classification problems
- B. softmax classifier is designed for multiclass classification problems
- C. softmax classifier can be interpreted by maximum likelihood estimation
- D. logistic regression cannot be interpreted by maximum likelihood estimation
- E. we can use stochastic gradient descent to solve both logistic regression and softmax classifier problems.

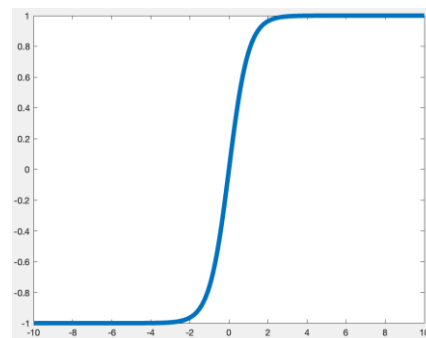
Your answer: AE

3. (multiple choice question, 5 points) which of the following statements are true?

- A. if we only use convex activation layers, we will have a convex neural network model
- B. nonlinearity is due to the nonlinear activation layers
- C. deep neural networks allow us to learn features, so we do not require to design hand crafted features
- D. If we use only linear functions in all layers (including activation layers), we will have a linear model.
- E. both ReLU and Tanh are differentiable.



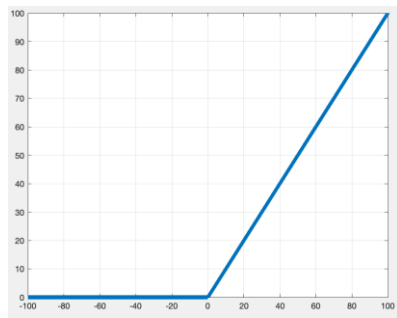
ReLU



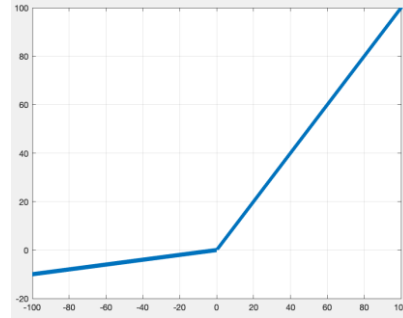
Tanh

Your answer: ABC

4. (single choice question + explanation, 6 points) Which of the following activation functions is more likely to prevent vanishing gradients? Explain why (up to 3 sentences).



A. ReLU



B. leaky ReLU

Your answer: B, B is the optimization of A.

5. (short answer, up to 5 sentences, 8 points) The input image has been converted into a matrix of size 28 X 28 and we have a convolutional kernel/filter of size 7 X 7 with a stride of 1. What will be the size of the output convoluted matrix? If we change the stride to 3 (for both horizontal and vertical directions), what will be the size of the convoluted matrix?

Your answer: $28-7+1 = 22$; $28-7+3 = 24$

6. (6 points) If we use different initialization points of the convolutional neural network model for stochastic gradient descent optimization algorithm:

(1) (single choice question) which of the following statements are true?

- A. it guarantees to converge to the same stationary point
- B. it cannot guarantee to converge to the same stationary point.

(2) (single choice question) the reason for your answer to question (1) is

- A. stochastic gradient descent is a commonly used and powerful algorithm.
- B. there is usually no closed form solution
- B. the network model contains the composition structure of functions
- C. the network model is nonlinear

Your answer:

(1)

B

(2)

A

7. (multiple choice question, 5 points) You are designing a house price predictor with hand crafted features. You can generate bag-of-words features for

- A. text description of the house
- B. image of the house
- C. zip code of the house
- D. the number of bedrooms
- E. land size of the house

Your answer: BDE

8. (short answer, up to 5 sentences, 8 points) Given the following input matrix and convolutional filter, what value would be in place of the question mark in output convoluted matrix? What is the stride?

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

input

1	0	1
0	1	1
1	0	1

conv filter/kernel

	?	

convoluted matrix

Your answer: stride is 1 because $5(\text{input}) - 3(\text{conv filter/kernel}) + 1(\text{stride}) = 3(\text{convoluted matrix})$ and ? in convoluted matrix is 2 because:

Input is

1	1	0
0	0	0
1	1	0

9. (short answer, up to 5 sentences, 8 points) In convolutional neural networks, does max pooling layers always decrease the parameters? Why? (hint: consider changing stride and pooling size)

Your answer: No, though the main function of the maximum pooling layer is to compress the image, but when the pool size is 1, the parameters are unchanged.

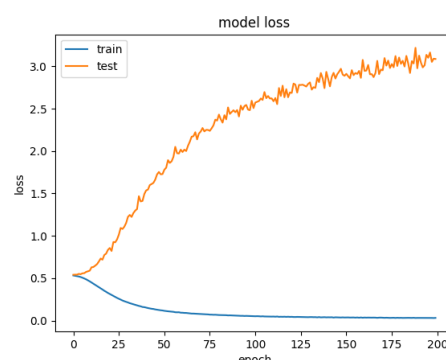
10. (multiple choice question, 5 points) Which of the following statements for back propagation (BP) are true?

- A. BP cannot be used to optimize the neural network if there is a max pooling layer
- B. BP is motivated by chain rule
- C. during BP, the computation of gradients for different layers can usually be performed in parallel
- D. forward computation is usually performed before BP can be computed
- E. BP may cause vanishing gradients.

Your answer: BDE

11. (multiple choice question + explanation, 6 points) You trained a convolutional neural network model and plot the training and testing loss. Explain why (up to 5 sentences).

- A. add dropout layers
- B. use a larger learning rate (step size)
- C. use ReLU function for activation layers
- D. add regularization
- E. perform data augmentation



Your answer: ADE

12. (multiple choice question, 5 points) Which of the following statements are true about a convolutional layer?

- A. the number of parameters in a filter depends on the number of channels of the input feature map.
- B. the number of channels of output feature map is equal to the number of filters.
- C. the total number of parameters in a filter depends on the stride.
- D. convolutional operation alone will decrease the dimension of feature map
- E. the number of filters depends on the number of input images in a batch

Your answer: CD

13. (short answer, up to 5 sentences for each question, 10 points) Briefly explain:

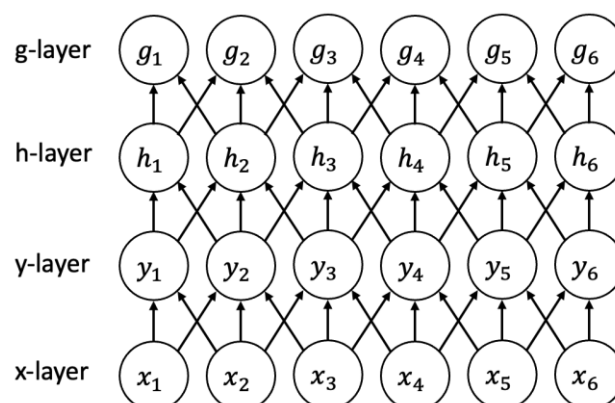
(1) why a convolutional neural network may require a pre-defined resolution for input images?

Your answer: The standardization of the input features is conducive to improving the learning efficiency and performance of the convolutional neural network.

(2) how to resolve this resolution issue?

Your answer: Use gradient descent algorithm for learning.

14. (short answer, up to 5 sentences, 8 points) What are the receptive fields of g_3 in h-layer, y-layer and x-layer, respectively?



Your answer: h-layer: h2,h3,h4

y-layer: y1,y2,y3,y4,y5

x-layer: 0,x1,x2,x3,x4,x5,x6

15. (short answer, up to 5 sentences, 9 points) You trained a convolutional neural network model for classifying 10 different brands of cars. Now you collect some images of a new brand and would like to include this new brand as the 11th class in your classifier. What technique would you use? Briefly explain your idea.

Your answer: Use Transfer learning, by using pre-trained models that have been trained on large data sets before, we can directly use the corresponding structures and weights to apply them to the problems we are facing.