

X


<https://swayam.gov.in>

https://swayam.gov.in/nc_details/NPTEL

2111cs010077@mallareddyuniversity.ac.in ✓

NPTEL (<https://swayam.gov.in/explorer?ncCode=NPTEL>) » Deep Learning - IIT Ropar (course)

Course outline

About NPTEL ()

How does an NPTEL online course work? ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 5 ()

Week 6 ()

Week 7 ()

☐ Bias and Variance (unit? unit=92&lesson=93)

☐ Train error vs Test error (unit?)

Week 7 : Assignment 7

The due date for submitting this assignment has passed.

Due on 2024-09-11, 23:59 IST.

Assignment submitted on 2024-09-11, 20:17 IST

Common Data Q1-Q2

Consider two models:

$$\hat{f}_1(x) = w_0 + w_1x$$

$$\hat{f}_2(x) = w_0 + w_1x^2 + w_2x^2 + w_4x^4 + w_5x^5$$

1) Which of these models has higher complexity?

1 point

☐ $\hat{f}_1(x)$

☒ $\hat{f}_2(x)$

☐ It is not possible to decide without knowing the true distribution of data points in the dataset.

Yes, the answer is correct.

Score: 1

Accepted Answers:

$$\hat{f}_2(x)$$

2) We generate the data using the following model:

1 point

$$y = 5x^3 + 2x + x + 3.$$

We fit the two models $\hat{f}_1(x)$ and $\hat{f}_2(x)$ on this data and train them using a neural network.

unit=92&lesso
n=94)

☐ Train error vs
Test error
(Recap) (unit?
unit=92&lesso
n=95)

☐ True error and
Model
complexity
(unit?
unit=92&lesso
n=96)

☐ L2
regularization
(unit?
unit=92&lesso
n=97)

☐ Dataset
augmentation
(unit?
unit=92&lesso
n=98)

☐ Parameter
sharing and
tying (unit?
unit=92&lesso
n=99)

☐ Adding Noise
to the inputs
(unit?
unit=92&lesso
n=100)

☐ Adding Noise
to the outputs
(unit?
unit=92&lesso
n=101)

☐ Early stopping
(unit?
unit=92&lesso
n=102)

☐ Ensemble
Methods (unit?
unit=92&lesso
n=103)

☐ Dropout (unit?
unit=92&lesso
n=104)



$\hat{f}_1(x)$ has a higher bias than $\hat{f}_2(x)$.



$\hat{f}_1(x)$ has a higher variance than $\hat{f}_2(x)$.



$\hat{f}_2(x)$ has a higher bias than $\hat{f}_1(x)$.



$\hat{f}_2(x)$ has a higher variance than $\hat{f}_1(x)$.

Yes, the answer is correct.

Score: 1

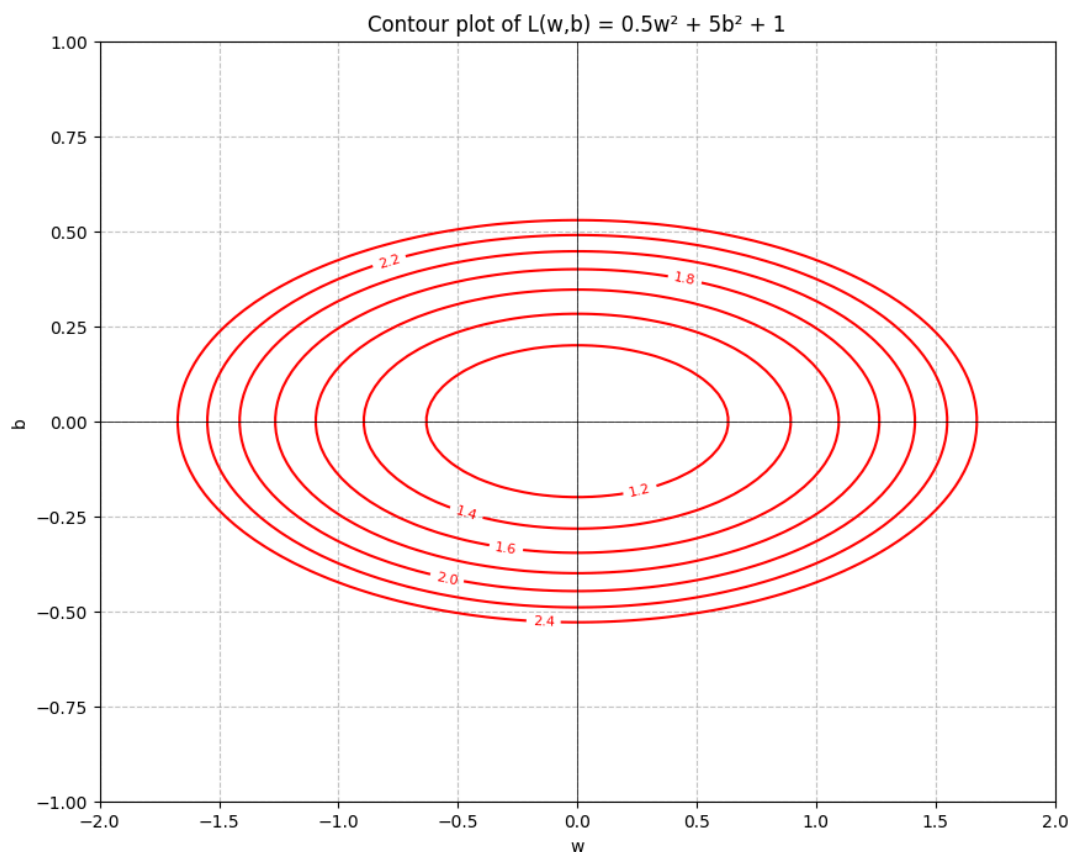
Accepted Answers:

$\hat{f}_1(x)$ has a higher bias than $\hat{f}_2(x)$.

$\hat{f}_2(x)$ has a higher variance than $\hat{f}_1(x)$.

Common Data Q3-Q6

Consider a function $L(w, b) = 0.5w^2 + 5b^2 + 1$ and its contour plot given below:



3) What is the value of $L(w^*, b^*)$ where w^* and b^* are the values that minimize the function.

1

Yes, the answer is correct.

Score: 1

Accepted Answers:

(Type: Range) 0.9, 1.1

1 point

4) What is the sum of the elements of $\nabla L(w^*, b^*)$?

☐ Lecture
Material for
Week 7 (unit?
unit=92&lesso
n=105)

☒ **Quiz: Week 7
: Assignment
7
(assessment?
name=295)**

☐ Week 7
Feedback
Form: Deep
Learning - IIT
Ropar (unit?
unit=92&lesso
n=236)

Week 8 ()

Week 9 ()

week 10 ()

Week 11 ()

Week 12 ()

**Download
Videos ()**

Books ()

**Text
Transcripts
()**

**Problem
Solving
Session -
July 2024 ()**

0

Yes, the answer is correct.

Score: 1

Accepted Answers:

(Type: Numeric) 0

1 point

5) What is the determinant of $H_L(w^*, b^*)$, where H is the Hessian of the function?

10

Yes, the answer is correct.

Score: 1

Accepted Answers:

(Type: Numeric) 10

1 point

6) Compute the Eigenvalues and Eigenvectors of the Hessian. According to the eigen- **1 point**
values of the Hessian, which parameter is the loss more sensitive to?

☒
 b

☐
 w

Yes, the answer is correct.

Score: 1

Accepted Answers:

b

7) Suppose that a model produces zero training error. What happens if we use L_2 **1 point**
regularization, in general?

- ☒ It might increase training error
- ☒ It might decrease test error
- ☐ It might decrease training error
- ☒ Reduce the complexity of the model by driving less important weights to close to zero

Yes, the answer is correct.

Score: 1

Accepted Answers:

It might increase training error

It might decrease test error

Reduce the complexity of the model by driving less important weights to close to zero

8) Suppose that we apply Dropout regularization to a feed forward neural network. **1 point**
Suppose further that mini-batch gradient descent algorithm is used for updating the parameters
of the network. Choose the correct statement(s) from the following statements.

- ☒ The dropout probability p can be different for each hidden layer
- ☐ Batch gradient descent cannot be used to update the parameters of the network
- ☒ Dropout with $p = 0.5$ acts as a ensemble regularize
- ☐ The weights of the neurons which were dropped during the forward propagation at t^{th}

iteration will not get updated during $t + 1^{th}$ iteration

Yes, the answer is correct.

Score: 1

Accepted Answers:

The dropout probability p can be different for each hidden layer

Dropout with $p = 0.5$ acts as a ensemble regularize

9) We have trained four different models on the same dataset using various hyperparameters. The training and validation errors for each model are provided below. Based on this information, which model is likely to perform best on the test dataset? **1 point**

Model	Training error	Validation error
1	0.9	1.2
2	0.3	0.6
3	1.5	0.5
4	1.2	1.2

☐ Model 1

☒ Model 2

☐ Model 3

☐ Model 4

Yes, the answer is correct.

Score: 1

Accepted Answers:

Model 2

10) Consider the problem of recognizing an alphabet (in upper case or lower case) of English language in an image. There are 26 alphabets in the language. Therefore, a team decided to use CNN network to solve this problem. Suppose that data augmentation technique is being used for regularization. Then which of the following transformation(s) on all the training images is (are) appropriate to the problem **1 point**



Rotating the images by $\pm 10^\circ$



Rotating the images by $\pm 180^\circ$



Translating image by 1 pixel in all direction



Cropping

Yes, the answer is correct.

Score: 1

Accepted Answers:

Rotating the images by $\pm 10^\circ$

Translating image by 1 pixel in all direction

Cropping