X

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

shweta_s@cs.iitr.ac.in

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

Click to register for Certification exam (https://examform.nptel.ac.in/2025_10/exam_form/dashboard)

If already registered, click to check your payment status

Course outline

About NPTEL ()

How does an NPTEL online course work? ()

Week 0 ()

Week 1 ()

Week 2 ()

Week 3 ()

week 4 ()

Week 7 : Assignment 7

The due date for submitting this assignment has passed.

Due on 2025-03-12, 23:59 IST.

Assignment submitted on 2025-03-12, 19:09 IST

1) Which of the following statements about L2 regularization is true? ***1 point***

- It adds a penalty term to the loss function that is proportional to the absolute value of the weights
- It results in sparse solutions for w
- It adds a penalty term to the loss function that is proportional to the square of the weights
- It is equivalent to adding Gaussian noise to the weights

Yes, the answer is correct.
Score: 1

Accepted Answers:
It adds a penalty term to the loss function that is proportional to the square of the weights

Common Data Q2-Q3

Consider two models:

$$\hat{f}_1(x) = w_0 + w_1 x$$

$$\hat{f}_2(x) = w_0 + w_1 x^2 + w_2 x^2 + w_4 x^4 + w_5 x^5$$

2) 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

Week 5 ()

Week 6 ()

Week 7 ()

- Bias and Variance (unit?unit=92&lesson=93)
- Train error vs Test error (unit?unit=92&lesson=94)
- Train error vs Test error (Recap) (unit?unit=92&lesson=95)
- True error and Model complexity (unit?unit=92&lesson=96)
- L2 regularization (unit?unit=92&lesson=97)
- Dataset augmentation (unit?unit=92&lesson=98)
- Parameter sharing and tying (unit?unit=92&lesson=99)
- Adding Noise to the inputs (unit?unit=92&lesson=100)

Yes, the answer is correct.
Score: 1

Accepted Answers:
$\hat{f}_2(x)$

3) We generate the data using the following model: ***1 point***

$$y = 7x^3 + 12x + x + 2.$$

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

- [x] $\hat{f}_1(x)$ has a higher bias than $\hat{f}_2(x)$
- [ ] $\hat{f}_2(x)$ has a higher bias than $\hat{f}_1(x)$
- [x] $\hat{f}_2(x)$ has a higher variance than $\hat{f}_1(x)$
- [ ] $\hat{f}_1(x)$ has a higher variance than $\hat{f}_2(x)$

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* $\hat{f}_1(x)$

4) 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

No, the answer is incorrect.
Score: 0

Accepted Answers:
The dropout probability p can be different for each hidden layer
Dropout with $p = 0.5$ acts as a ensemble regularize

5) 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***

- Adding Noise to the outputs (unit?unit=92&lesson=101)
- Early stopping (unit?unit=92&lesson=102)
- Ensemble Methods (unit?unit=92&lesson=103)
- Dropout (unit?unit=92&lesson=104)
- Lecture Material for Week 7 (unit?unit=92&lesson=105)
- Week 7 Feedback Form:Deep Learning - IIT Ropar!! (unit?unit=92&lesson=236)
- Week 7: Solution (unit?unit=92&lesson=251)
- **Quiz: Week 7 : Assignment 7 (assessment?name=316)**

Week 8 ()

Week 9 ()

week 10 ()

Week 11 ()

Week 12 ()

Download Videos ()

Model	Training error	Validation error
1	0.8	1.4
2	2.5	0.5
3	1.7	1.7
4	0.2	0.6

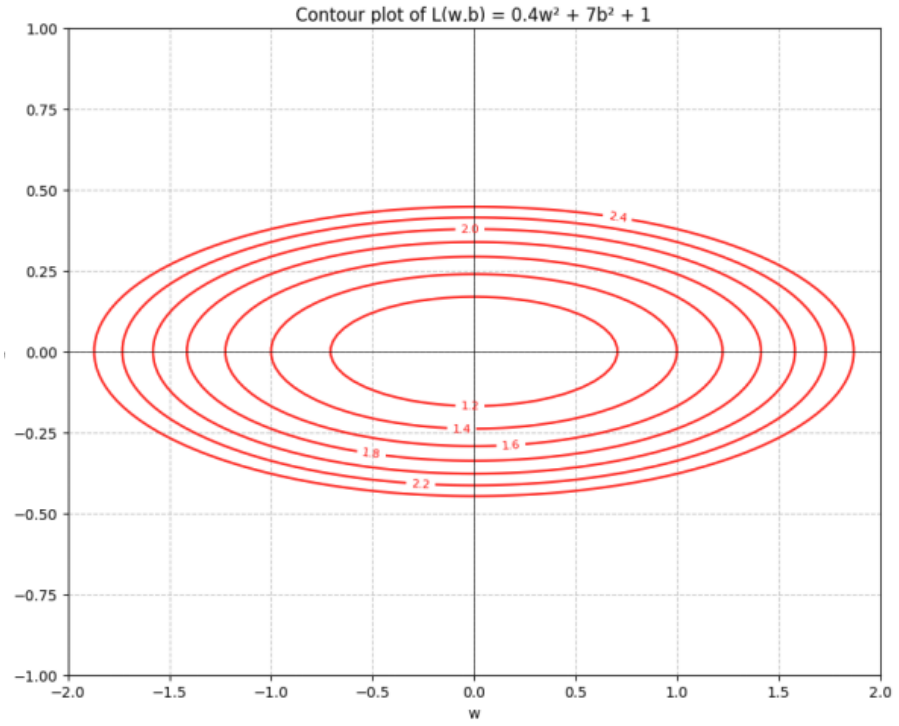
- Model 1
- Model 2
- Model 3
- Model 4

No, the answer is incorrect.
Score: 0

Accepted Answers:
Model 4

Common Data Q6-Q9

Consider a function $L(w, b) = 0.4w^2 + 7b^2 + 1$ and its contour plot given below:



6) 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: Numeric) 1

1 point

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

Books ()

Text Transcripts ()

Problem Solving Session - Jan 2025 ()

1

No, the answer is incorrect.
Score: 0

Accepted Answers:
(Type: Numeric) 0

1 point

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

11.2

Yes, the answer is correct.
Score: 1

Accepted Answers:
(Type: Numeric) 11.2

1 point

9) Compute the Eigenvalues and Eigenvectors of the Hessian. According to the eigenvalues of the Hessian, which parameter is the loss more sensitive to? ***1 point***

- b
- w

No, the answer is incorrect.
Score: 0

Accepted Answers:
b

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***

- [x] Rotating the images by $\pm 10^\circ$
- [ ] Rotating the images by $\pm 180^\circ$
- [x] Translating image by 1 pixel in all direction
- [ ] Cropping

Partially Correct.
Score: 0.66

Accepted Answers:
Rotating the images by $\pm 10^\circ$
Translating image by 1 pixel in all direction
Cropping