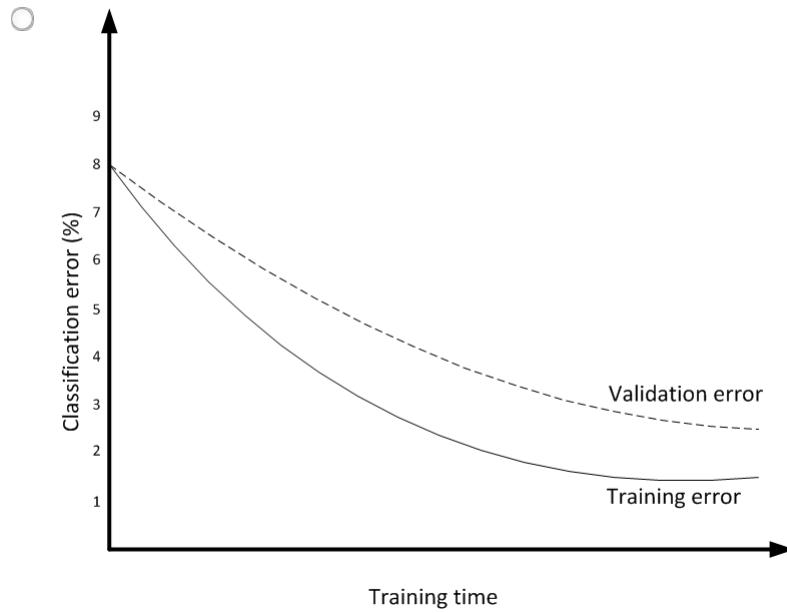
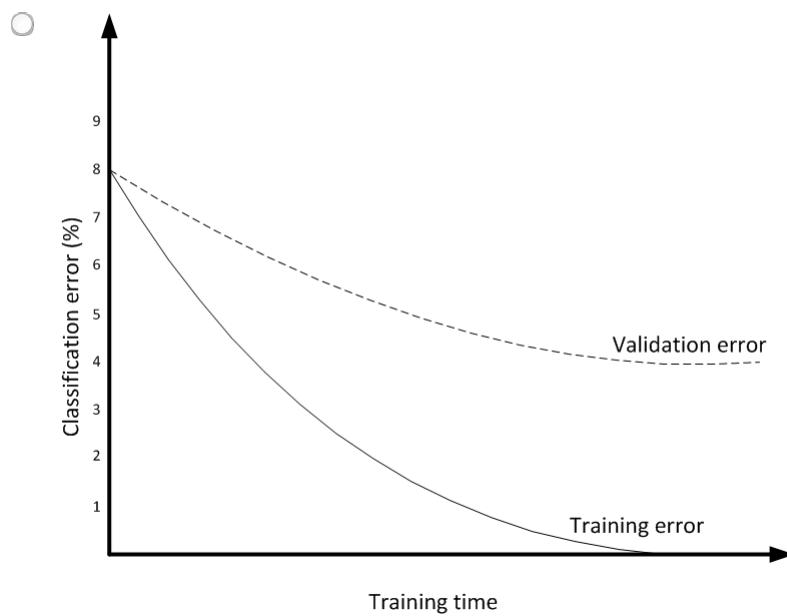


[←](#) Lecture 9 Quiz
Quiz, 6 questions

1 point

1.

You are experimenting with two different models for a classification task. The figures below show the classification error you get as training progresses on the training data and the validation data for each of the two models. Which model do you think would perform better on previously unseen test data?



1
point

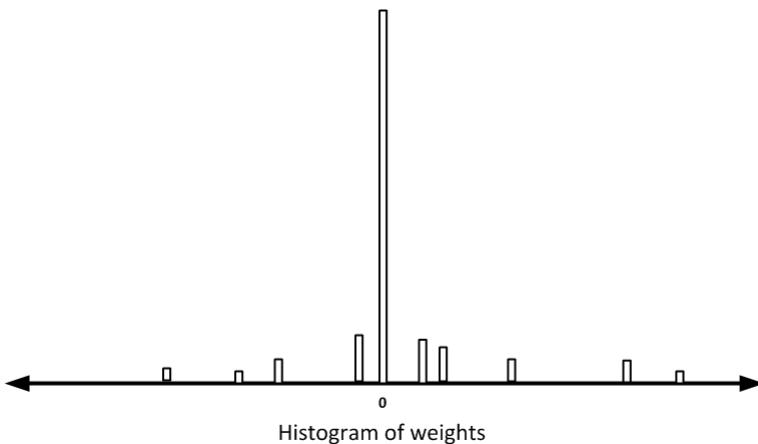


Lecture 9 Quiz

Quiz, 6 questions

2.

The figure below shows the histogram of weights for a learned Neural Network.



Which regularization technique has been used during learning?

- L2 regularization
 - adding weight noise
 - no regularization has been used
 - L1 regularization
-

1
point

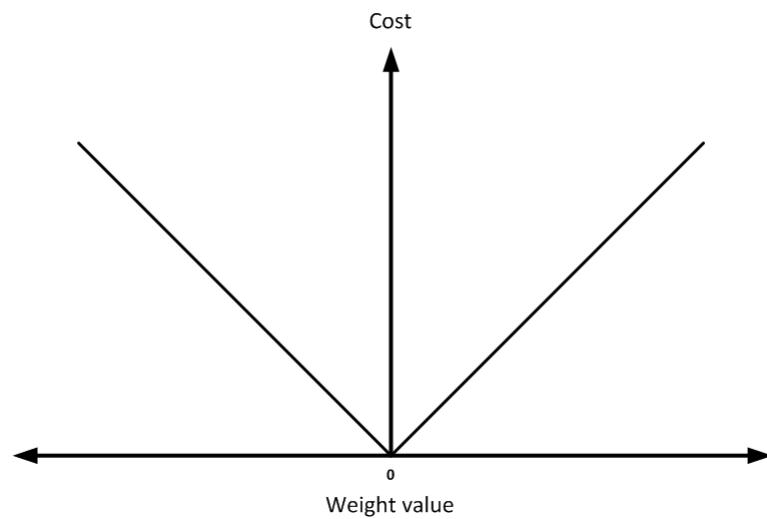
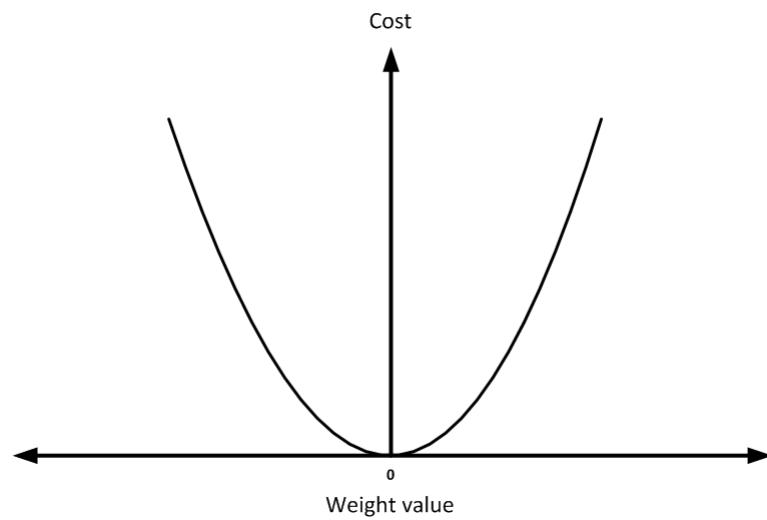
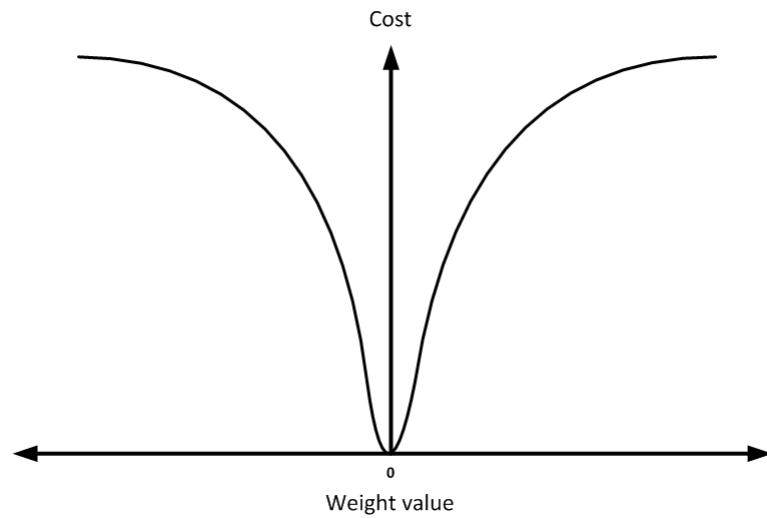


Lecture 9 Quiz

Quiz, 6 questions

3.

Suppose you want to regularize the weights of a neural network during training so that lots of its weights are quite close to zero, but a few are a very long way from zero. Which cost function you would add to your objective function?



1 point



Lecture 9 Quiz

Quiz, 6 questions

4.

In a linear regression task, a d dimensional input vector x is used to predict the output value y using the weight vector w where $y = w^T x$. The error function $E = \frac{1}{2}(t - w^T x)^2$ where t is the target output value. We want to use a student t cost for the weights: $C = \frac{\lambda}{2} \sum_{i=1}^d \log(1 + w_i^2)$.

The total error to be optimized $E_{tot} = E + C$. What is the expression for $\frac{\partial E_{tot}}{\partial w_i}$?

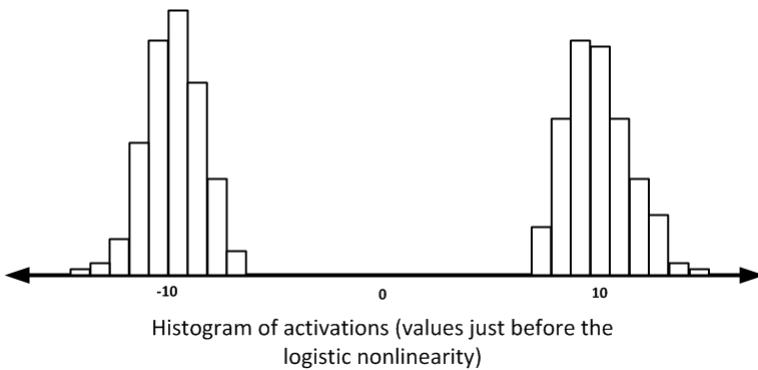
- $\frac{\partial E_{tot}}{\partial w_i} = -(t - w^T x)x_i + \frac{\lambda w_i}{(1+w_i^2)}$
- $\frac{\partial E_{tot}}{\partial w_i} = -(t - w_i x_i) - 2\lambda \frac{w_i}{(1+w_i^2)}$
- $\frac{\partial E_{tot}}{\partial w_i} = -(t - w^T x)x_i + \frac{\lambda}{(1+w_i^2)^2}$
- $\frac{\partial E_{tot}}{\partial w_i} = -(t - w_i x_i) - \lambda w_i$

1 point

5.

Different regularization methods have different effects on the learning process. For example *L2* regularization penalizes high weight values. *L1* regularization penalizes weight values that do not equal zero. Adding noise to the weights during learning ensures that the learned hidden representations take extreme values. Sampling the hidden representations regularizes the network by pushing the hidden representation to be binary during the forward pass which limits the modeling capacity of the network.

Given the shown histogram of activations (just before the nonlinear logistic nonlinearity) for a Neural Network, what is the regularization method that has been used (check all that apply)?



- Sampling the hidden representation
- L2* regularization
- Adding weight noise
- L1* regularization

1 point

6.

Suppose we have learned to predict a real-valued output from an input vector using a neural net with several hidden layers.

If we increase the amount of training data and train the network again, which of the following statements will probably be true:

- It will do better on the test data.
- It will do better on the training data.
- It will do worse on the training data.
- It will do worse on the test data.

I, **Aditya Manglik**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.
Learn more about Coursera's Honor Code

Submit Quiz

