Lecture 9 Quiz

3/6 points (50%)

Quiz, 6 questions

✖ Try again once you are ready.

Required to pass: 80% or higher

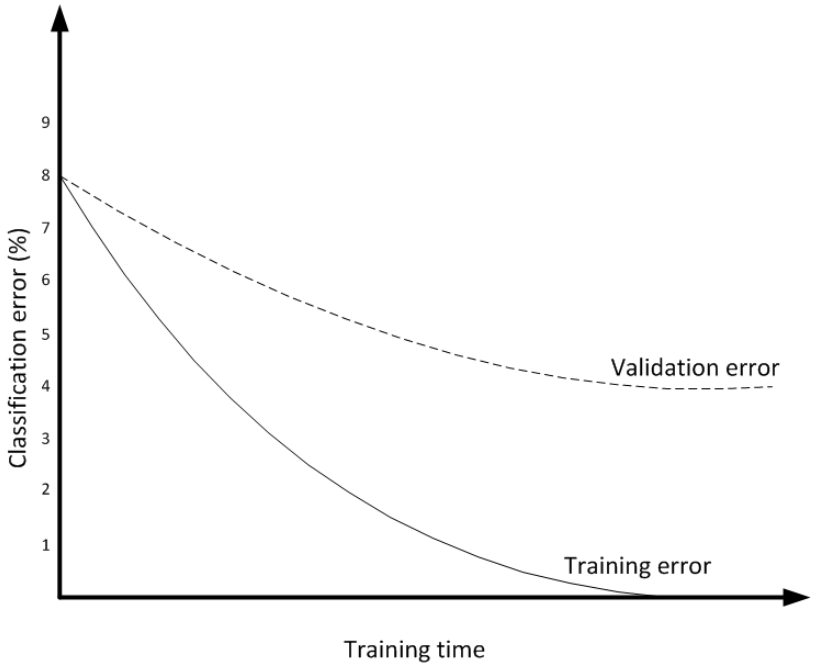
You can retake this quiz up to 3 times every 8 hours.

Back to Week 9

Retake

✖ 0 / 1 points

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?.

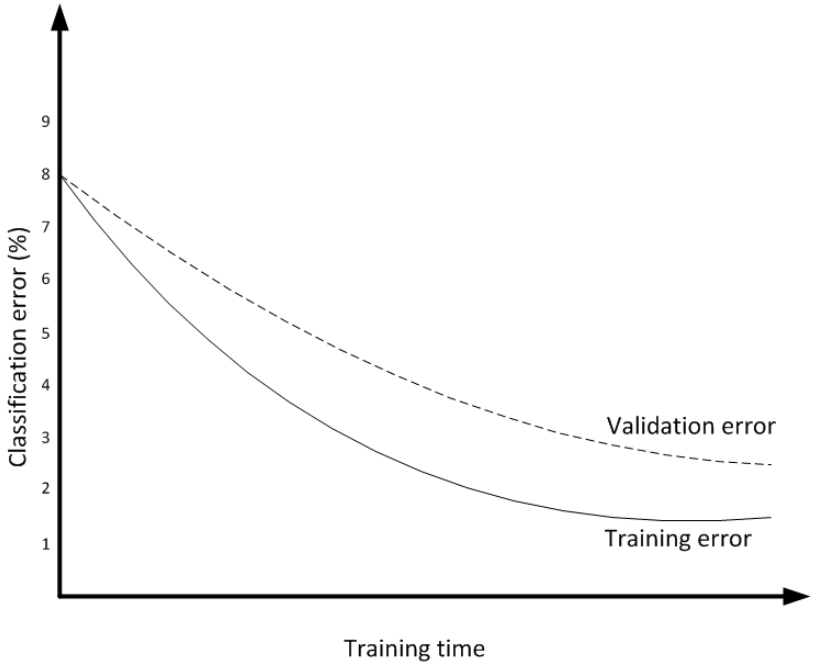


This should not be selected

Lecture 9 Quiz

Quiz, 6 questions

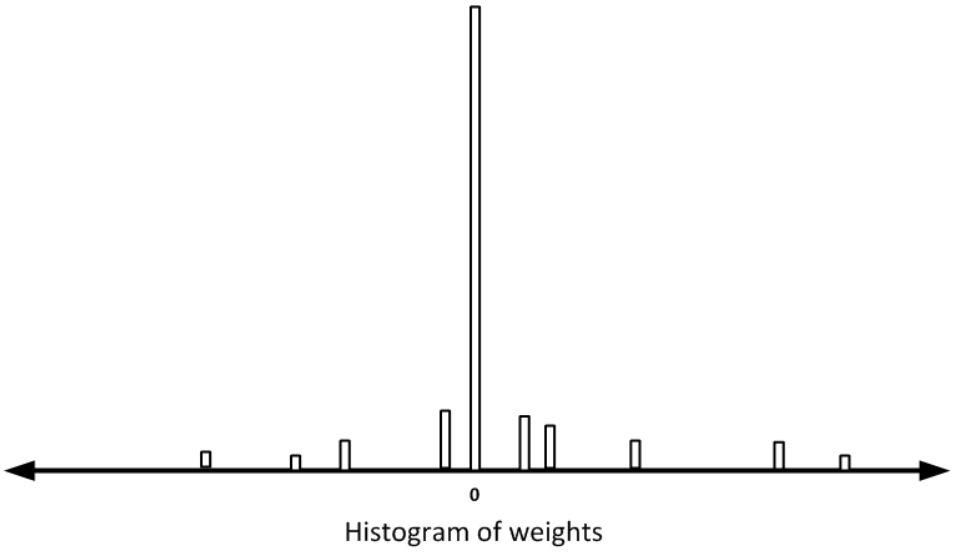
3/6 points (50%)



1 / 1 points

2.

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



Which regularization technique has been used during learning?

- no regularization has been used
- L1 regularization

Correct

Lecture 9 Quiz

Quiz, 6 questions

3/6 points (50%)

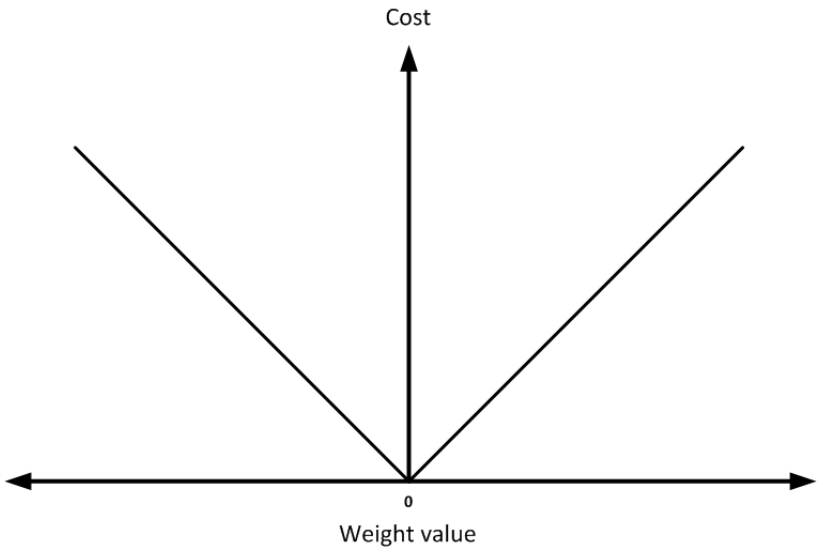
There are a lot of weights at zero. This is what you would expect with L1 regularization because it keeps pushing the weights towards zero with the same force no matter what size they are. Strong L2 regularization could also force a lot of weights to close to zero, but it would not allow those weights that are quite far from zero. Adding additive noise to the weights would not push them to zero (but adding multiplicative noise would).

- adding weight noise
- L2 regularization

✔ 1 / 1 points

3.

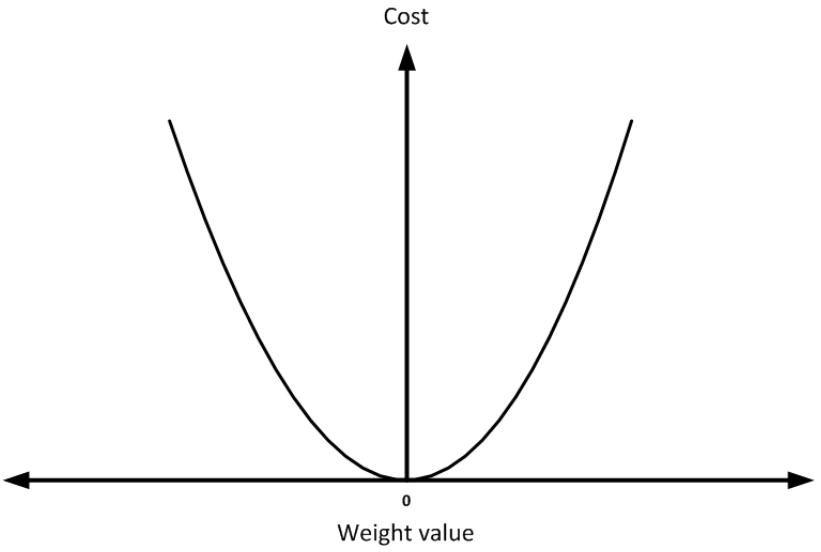
Suppose you want to regularize the weights of a neural network during training so that lots of its weights are quite close to zero, but a few are a very long way from zero. Which cost function you would add to your objective function?

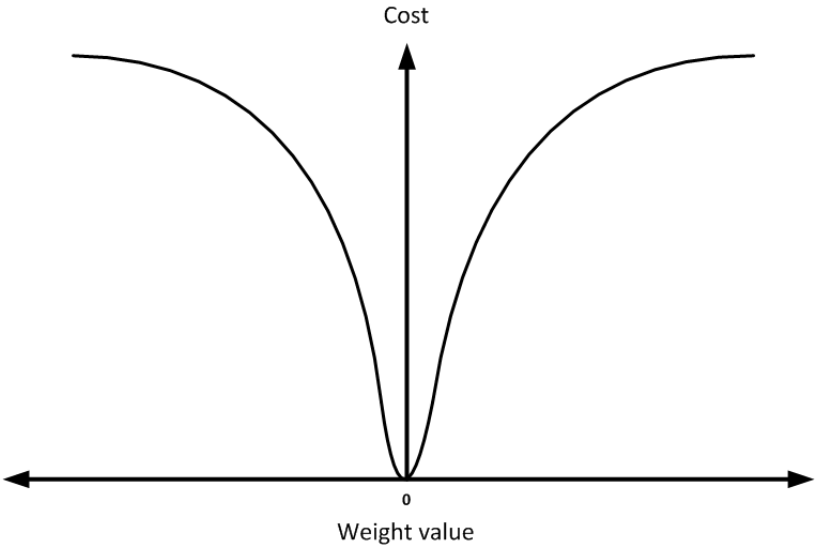


Lecture 9 Quiz

Quiz, 6 questions

3/6 points (50%)





Correct

L1 regularization would usually force many of the weights to be exactly zero. Strong L2 regularization could force a lot of weights to be close to zero, but it would then be extremely costly to have any very big weights.

0 / 1 points

4.

Lecture 9 Quiz

Quiz, 6 questions

3/6 points (50%)

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 $E = \frac{1}{2}\left(t - w^T x\right)^2$ 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 $\frac{\partial E_{tot}}{\partial w_i}$?

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

  This should not be selected

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

✔ 1 / 1 points

5.

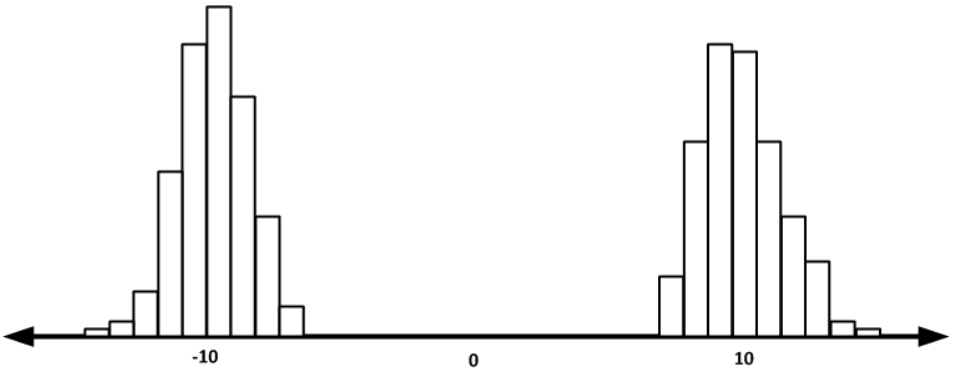
Lecture 9 Quiz

Quiz, 6 questions

3/6 points (50%)

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)?



Histogram of activations (values just before the logistic nonlinearity)

- [ ] $L1$ regularization

Un-selected is correct

- [ ] $L2$ regularization

Un-selected is correct

- [ ] Sampling the hidden representation

Correct

When you sample the hidden states, the sampling creates noise if the logistic is in its sensitive region. The learning tends to find solutions that minimize this noise by keeping units firmly on or firmly off.

- [ ] Adding weight noise

Correct