



Lecture 6 Quiz



5/5 points earned
(100%)

[Continue Course](#)[Back to Week 6](#)

Quiz passed!

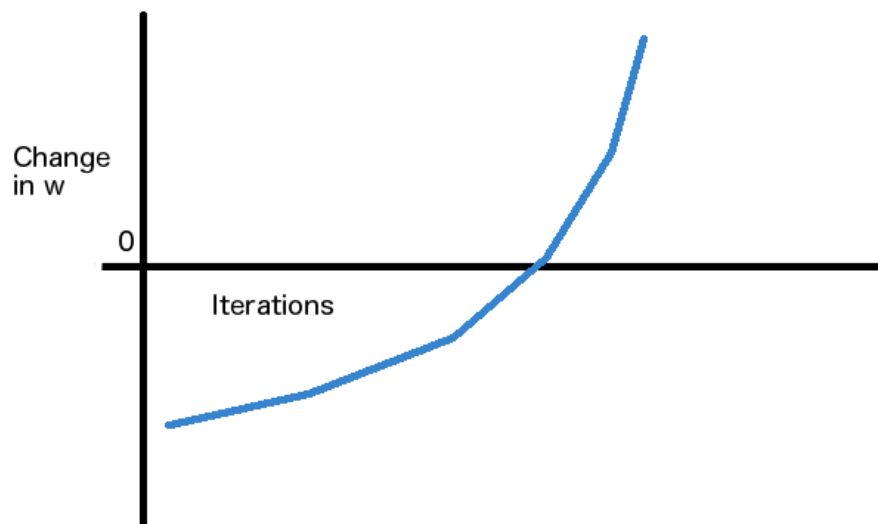


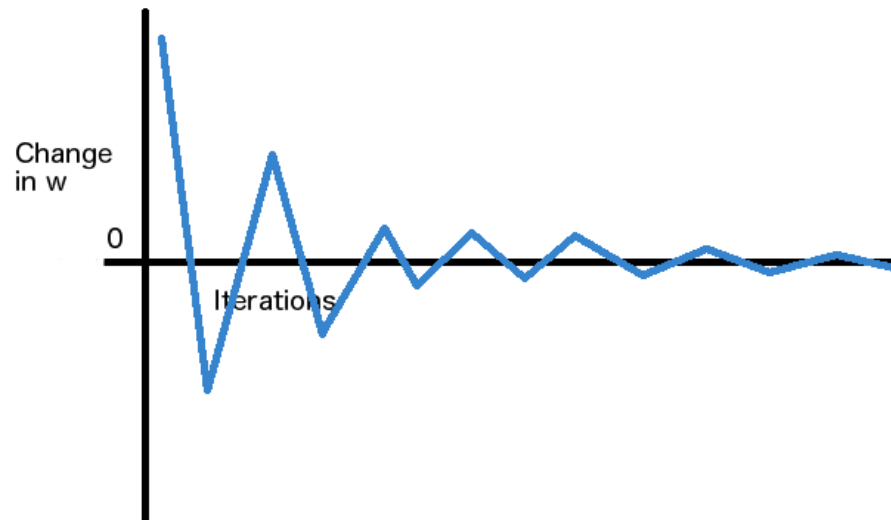
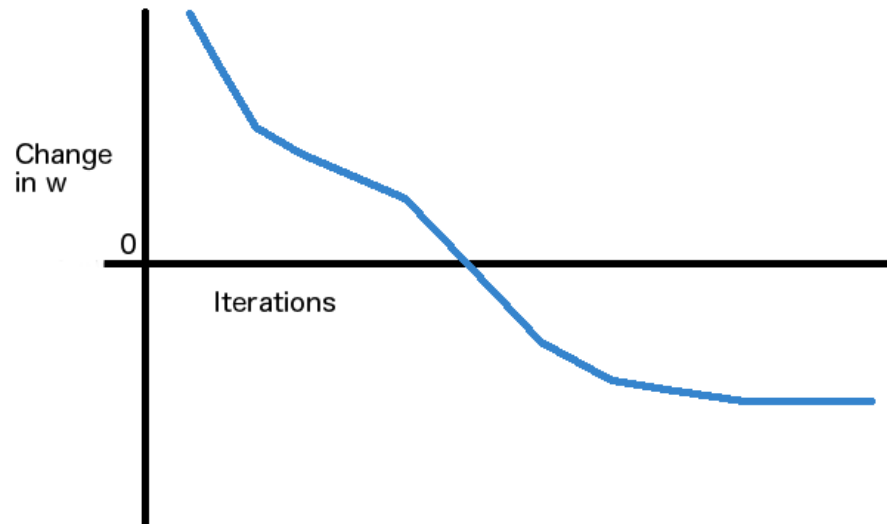
1 / 1
points

1.

Suppose w is the weight on some connection in a neural network. The network is trained using gradient descent until the learning *converges*. We plot the change of w as training progresses. Which of the following scenarios shows that convergence has occurred? **Notice that we're plotting the change in w , as opposed to w itself.**

Note that in the plots below, each *iteration* refers to a single *step* of steepest descent on a *single minibatch*.

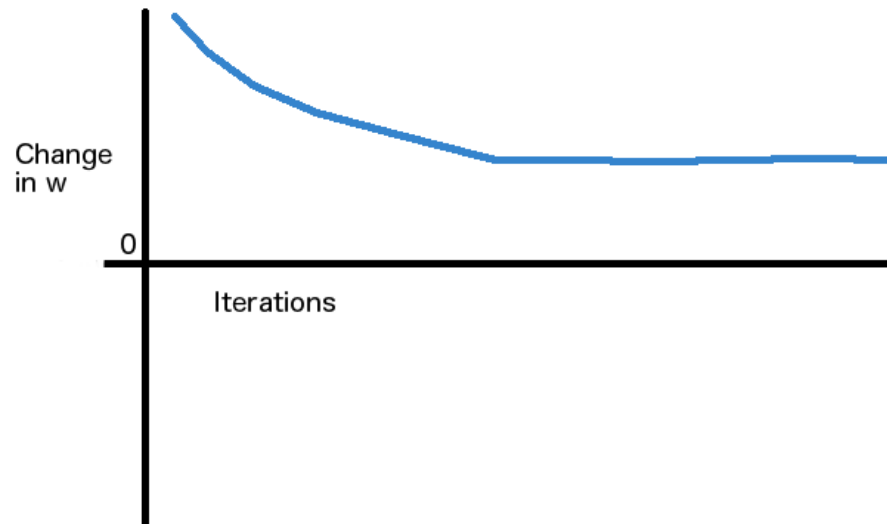




Correct

If the optimization has converged, w must converge to (or at most oscillate around) a point. So the change in w must converge to (or oscillate around) zero.





1 / 1
points

2.

Suppose you are using mini-batch gradient descent for training some neural nets on a large dataset. All neurons are logistic. You have to decide the mini-batch size and learning rate. You try some values and find that the value of the objective function on the training set keeps fluctuating and does not converge. What could be going wrong? Check all that apply.



The learning rate may be too big.



Correct

Large learning rates may cause divergent oscillations in weight space.



The size of the dataset may be too large.



Un-selected is correct



The weights may have been initialized to large values.



Un-selected is correct



The mini-batch size could be too small.



Correct

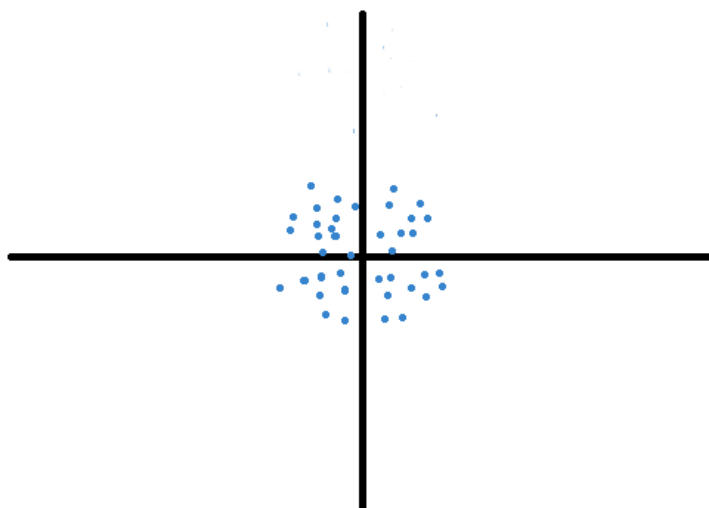
Small mini-batch may lead to more sampling noise in the estimated gradient.



1 / 1
points

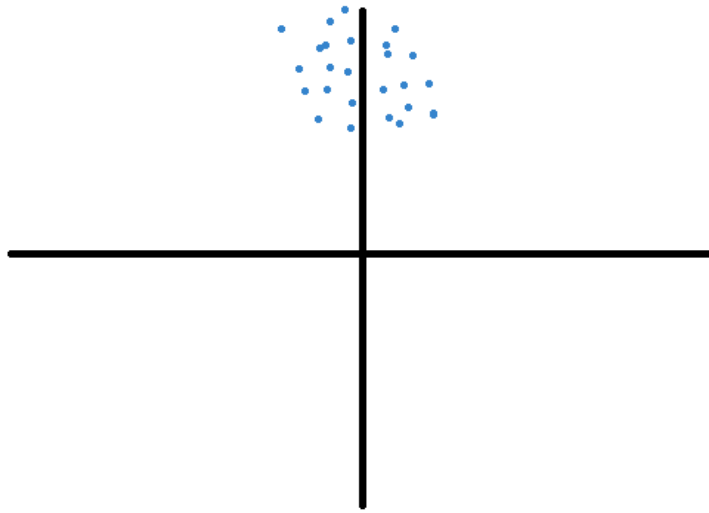
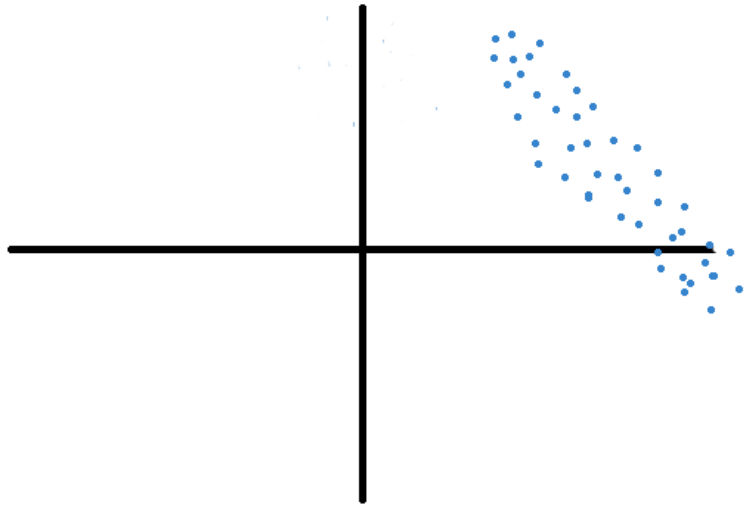
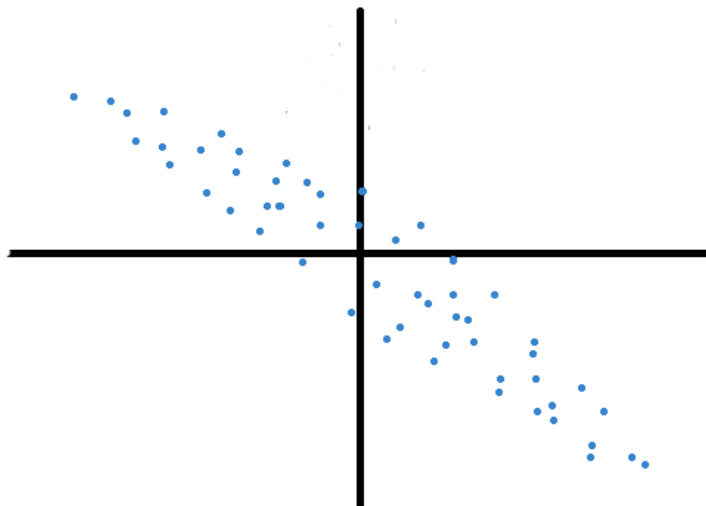
3.

Four datasets are shown below. Each dataset has two input values (plotted below) and a target value (not shown). Each point in the plots denotes one training case. Assume that we are solving a classification problem. Which of the following datasets would most likely be easiest to train using neural nets ?

**Correct**

This dataset has mean zero and diagonal covariance.

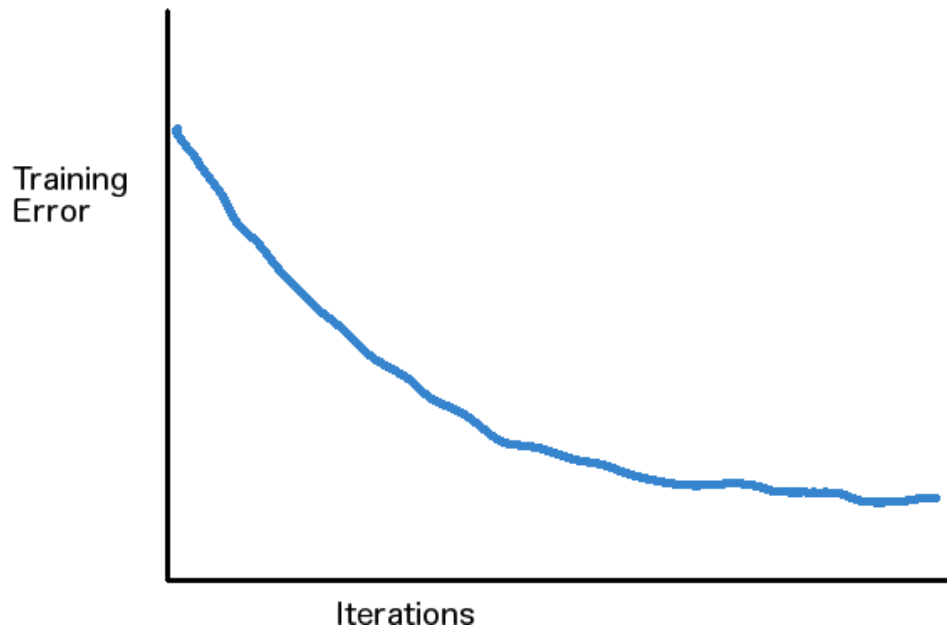


☐☐

1 / 1
points

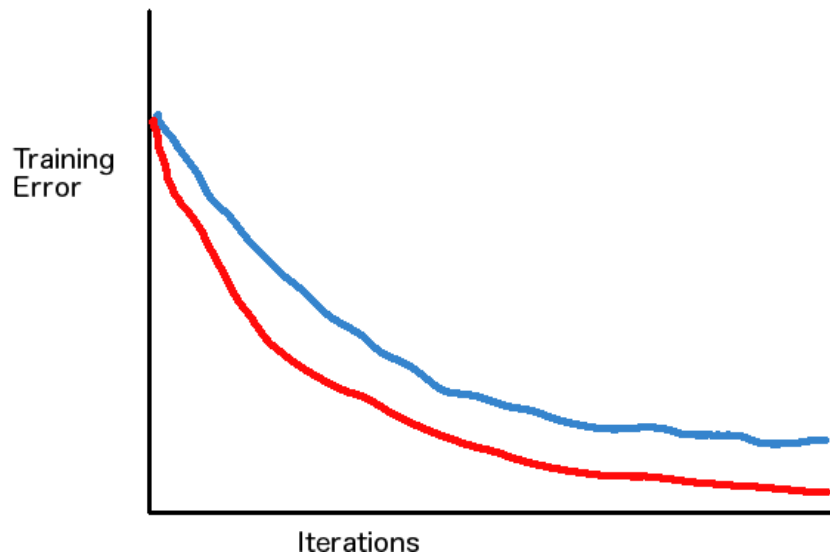
4.

Claire is training a neural net using mini-batch gradient descent. She chose a particular learning rate and found that the training error decreased as more iterations of training were performed as shown here in blue

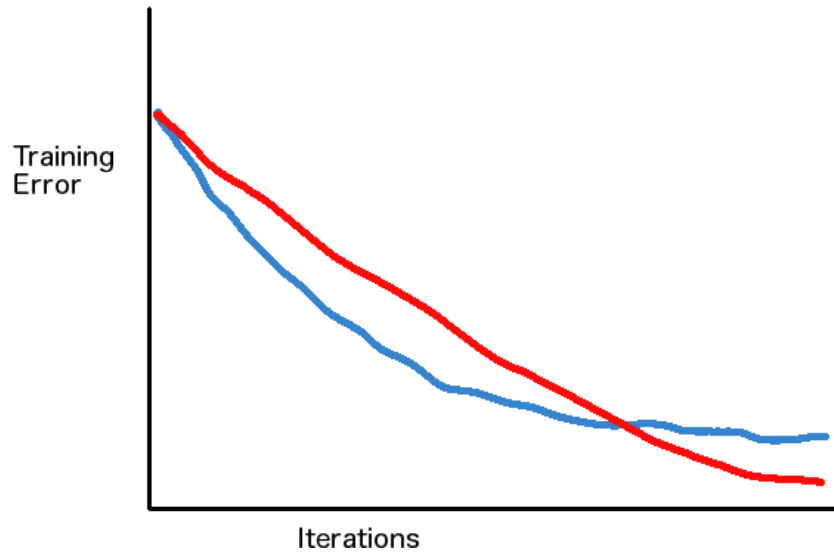


She was not sure if this was the best she could do. So she tried a **smaller** learning rate. Which of the following error curves (shown in red) might she observe now? Select the two most likely plots.

Note that in the plots below, each *iteration* refers to a single *step* of steepest descent on a *single minibatch*.

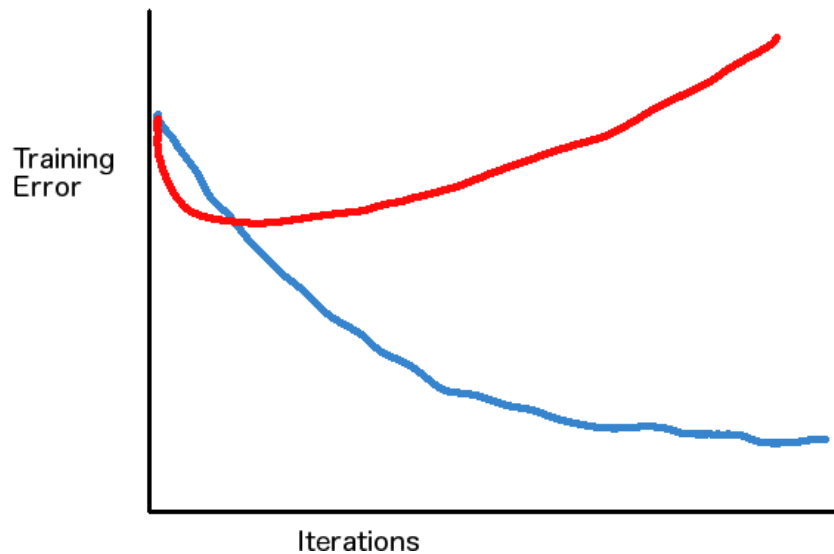


Un-selected is correct



Correct

A smaller learning rate may lead to slower progress initially but result in a lower final error.



Un-selected is correct

