Lecture 6 Quiz

Back to Week 6

4/5 points earned
(80%)

Quiz passed!

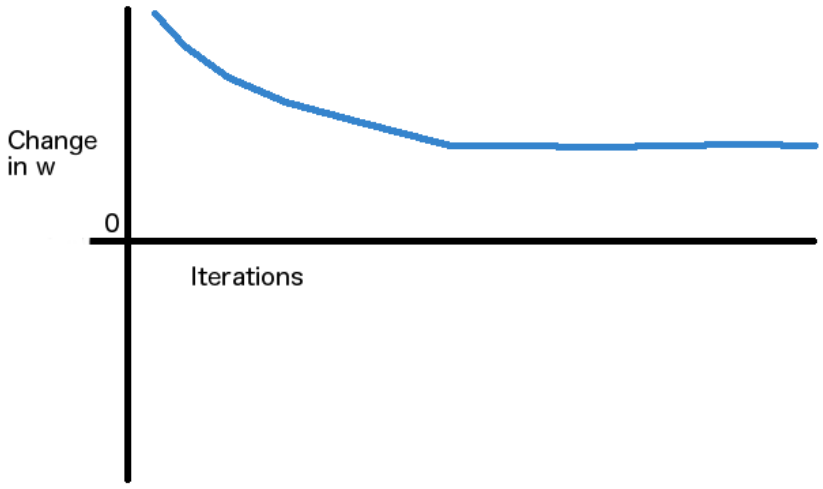
1 / 1
points

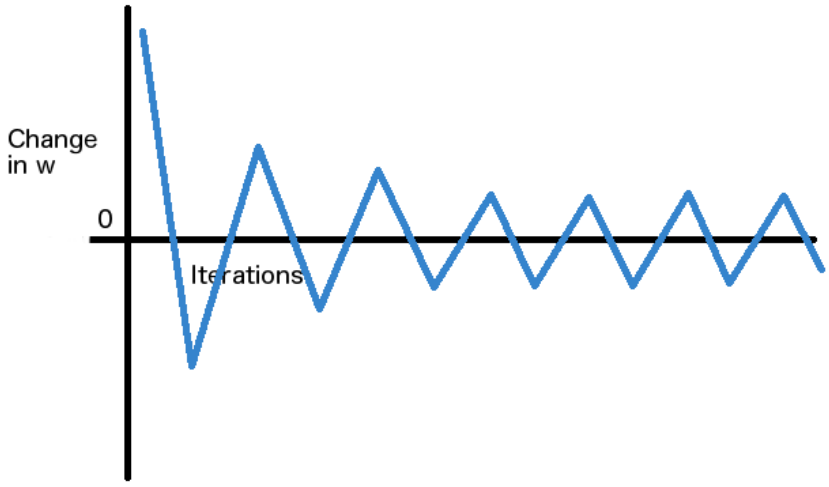
1.
Suppose w is the weight on some connection in a neural network. The network is trained using gradient descent until the learning

converges. However, the dataset consists of two mini-batches, which differ from each other somewhat. As usual, we alternate between the mini-batches for our gradient calculations, and that has implications for what happens

after convergence. We plot the change of w as training progresses. Which of the following scenarios shows that convergence has occurred? **Noticethat we're plotting the change in w, as opposed to w itself.**

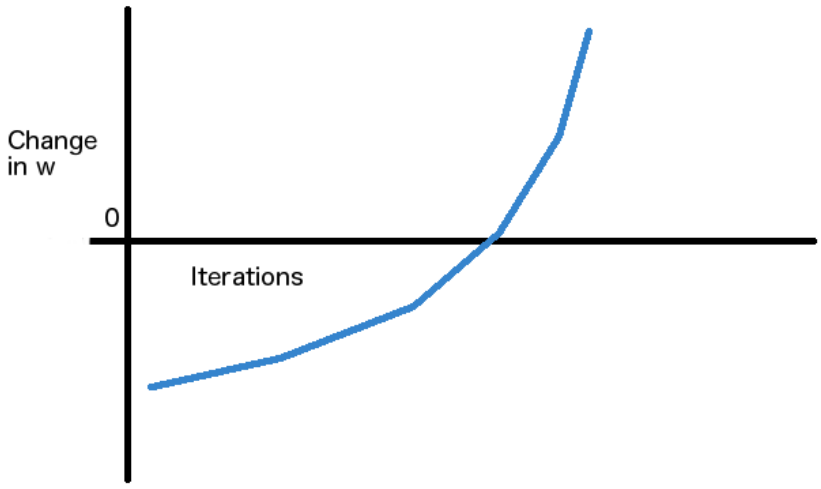
Note that in the plots below, each *iteration* refers to a single *step* of steepest descent on a *single minibatch*.



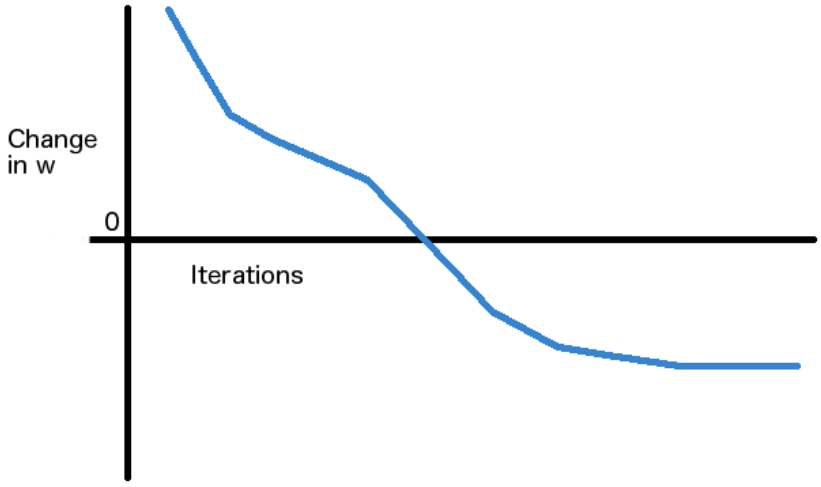


Correct

If the optimization has converged, w must converge to, or oscillate around, a point. So the change in w must converge to (or oscillate around) zero. The two mini-batches, however, will produce different gradients, one the opposite of the other. That causes w to go back and forth.



○



✖ 0 / 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 weights may have been initialized to large values.

This should not be selected

Large weights may lead to saturation of hidden units causing slow convergence ("plateauing") but not fluctuations if the output of each unit is bounded (as is the case with logistic units).

☐ The mini-batch size could be too small.

This should be selected

☐ The size of the dataset may be too large.

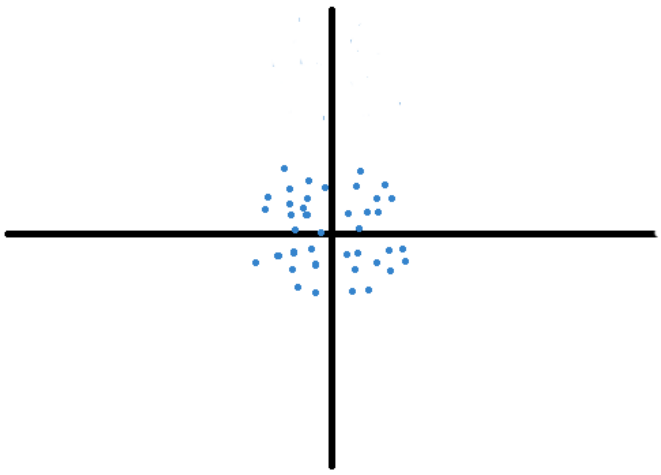
Un-selected is correct

✔ 1 / 1 points

3.

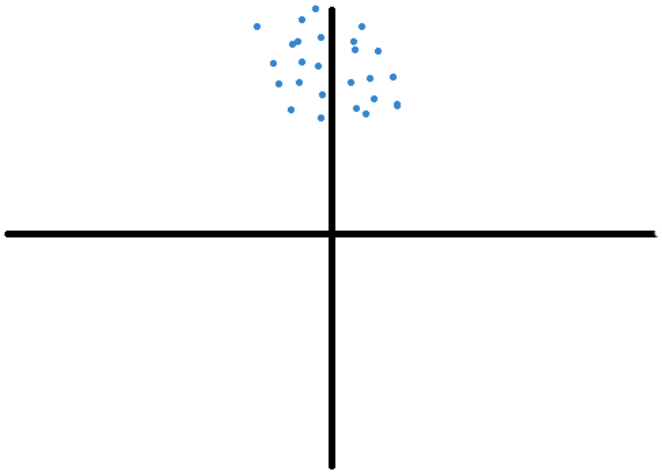
Four datasets are shown below. Each dataset has two input values (plotted below) and a target value (not shown). Each point in the plots denotes one training case. Assume that we are solving a classification problem. Which of the following datasets would most likely be easiest to train using neural nets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

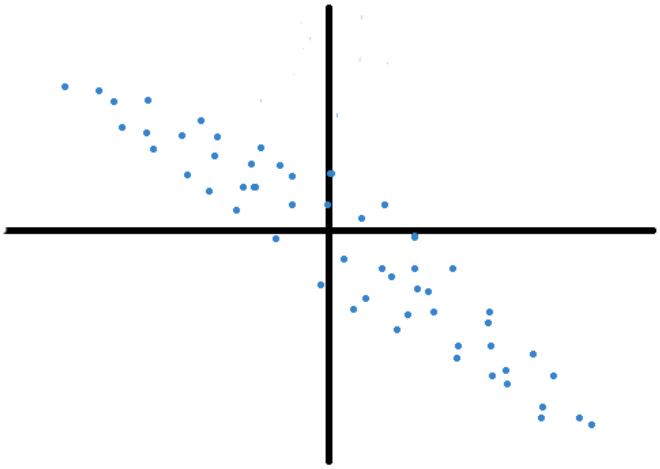
○

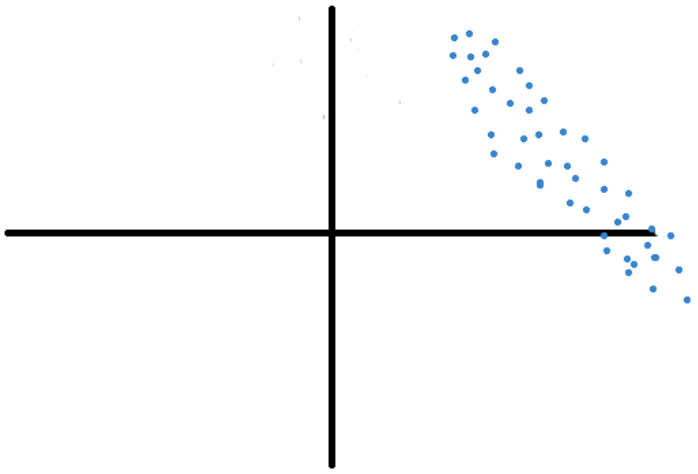


Correct

This dataset has mean zero and diagonal covariance.



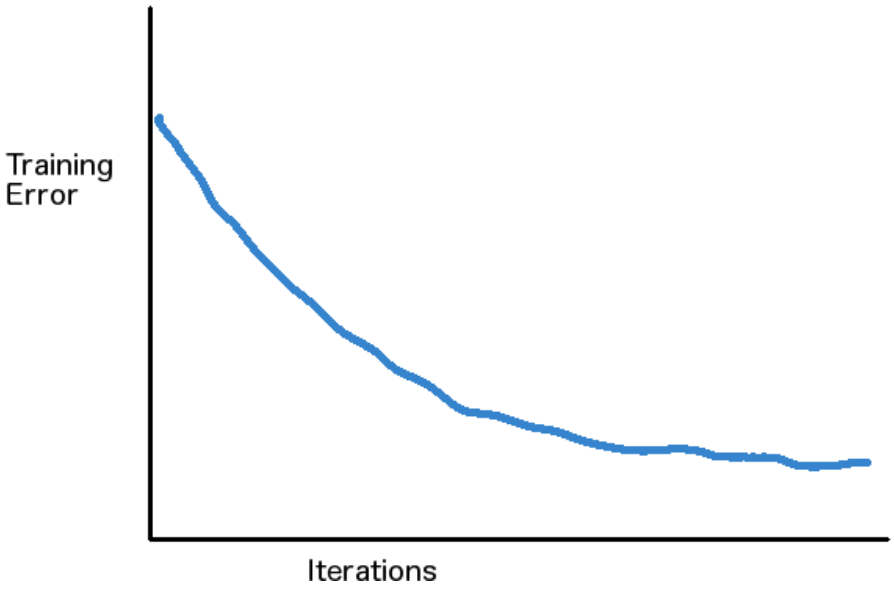




1 / 1 points

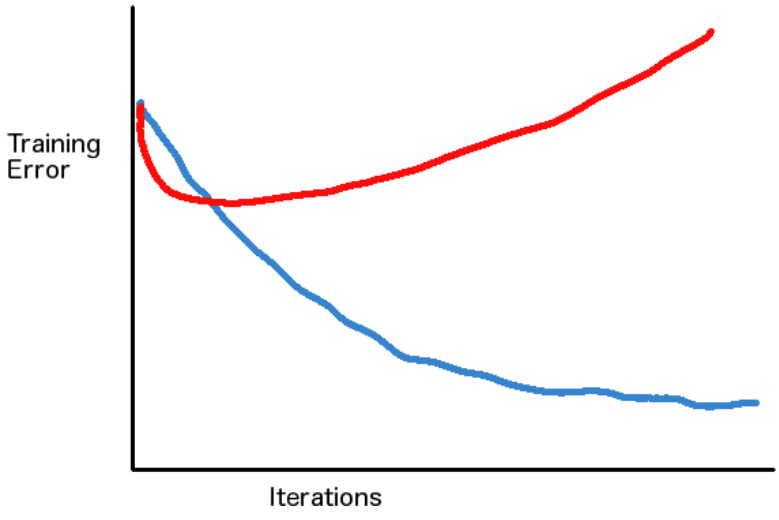
4.

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

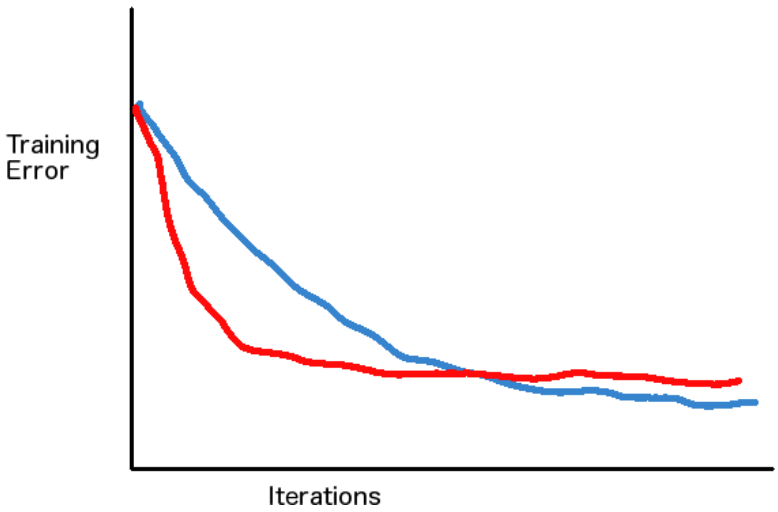


She was not sure if this was the best she could do. So she tried a **smaller** learning rate. Which of the following error curves (shown in red) might she observe now? Select the two most likely plots.

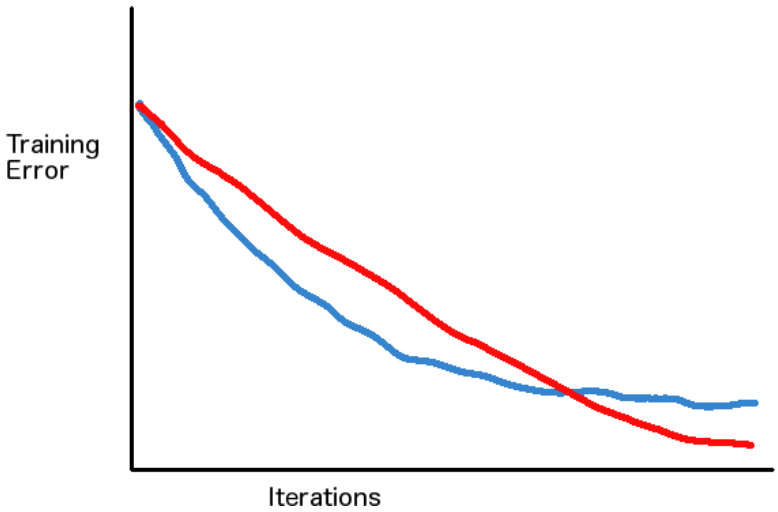
Note that in the plots below, each *iteration* refers to a single *step* of steepest descent on a *single minibatch*.



Un-selected is correct

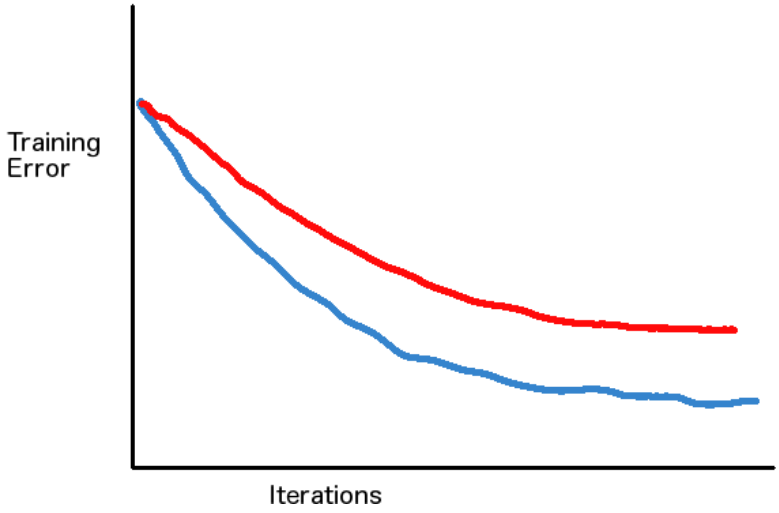


Un-selected is correct



Correct

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



Correct
A smaller learning rate may lead to slower convergence.

---

1 / 1 points

5.
In the lectures, we discussed two kinds of gradient descent algorithms: mini-batch and full-batch. For which of the following problems is mini-batch gradient descent likely to be **a lot better** than full-batch gradient descent?

- [ ] Sentiment Analysis: Decide whether a given movie review says that the movie is 'good' or 'bad'. The input consists of the word count in the review, for each of 50,000 words. The training set consists of 100 movie reviews written by experts for a newspaper.

Un-selected is correct

- [ ] Predict if an experiment at the Large Hadron Collider is going to yield positive results. The input consists of 25 experiment parameters (energy level, types of particles, etc). The training set consists of the 200 experiments that have already been completed (some of those yielded positive results; some yielded only negative results).

Un-selected is correct