Lecture 6 Quiz

5/5 points (100%)

Quiz, 5 questions

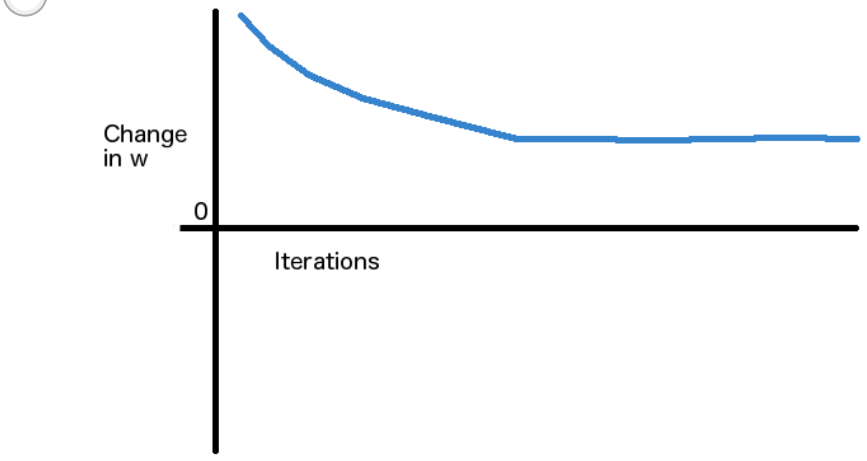
✔ Congratulations! You passed!

Next Item

✔ 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.**

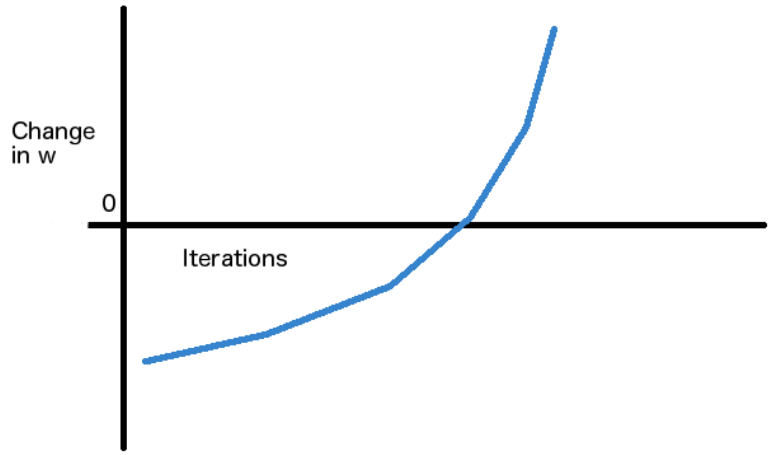
Note that in the plots below, each *iteration* refers to a single *step* of steepest descent on a *single minibatch*.

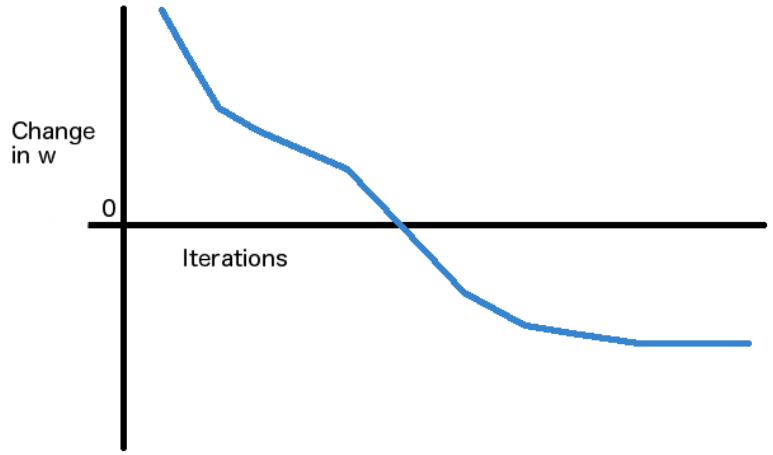


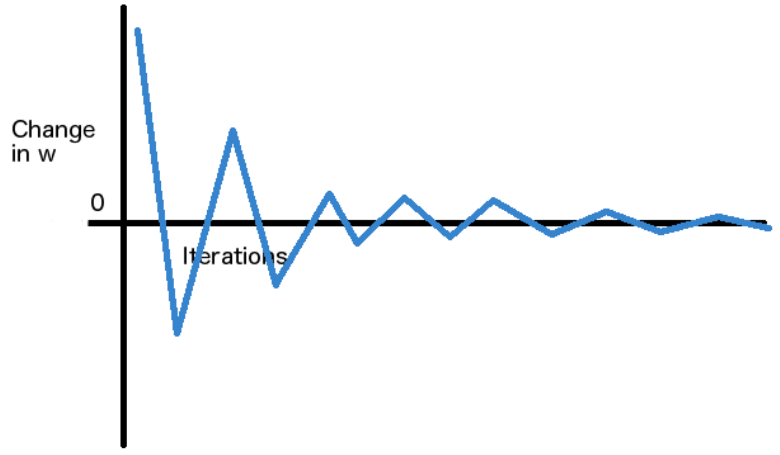
Lecture 6 Quiz

Quiz, 5 questions

5/5 points (100%)







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

Lecture 6 Quiz

Quiz, 5 questions

5/5 points (100%)

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.

- [x] The learning rate may be too big.

Correct
Large learning rates may cause divergent oscillations in weight space.

- [x] The mini-batch size could be too small.

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

- [ ] The weights may have been initialized to large values.

Un-selected is correct

- [ ] The size of the dataset may be too large.

Un-selected is correct

✔ 1 / 1 points

3.
Full-batch gradient descent can be used to minimize an objective function if the dataset is not too large. Which statement regarding full-batch gradient descent is **false**?

- For some setting of the learning rate, it is possible that the objective function increases in some iteration.

Lecture 6 Quiz

Quiz, 5 questions

5/5 points (100%)

Using momentum can be useful for full batch gradient descent.

Adaptive learning rate methods perform well for full-batch (or large mini-batch) gradient descent.

Full batch gradient descent is guaranteed to find a better local minimum than mini-batch gradient descent.

Correct

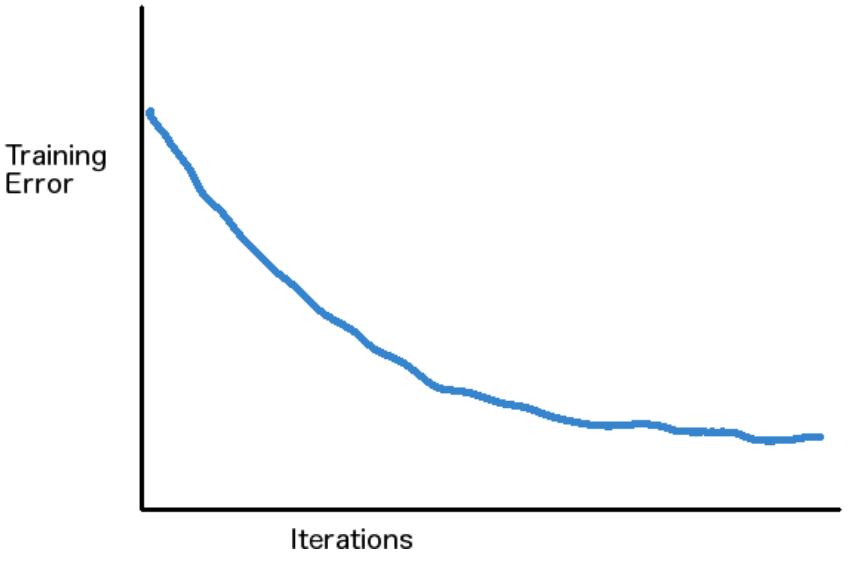
This is not necessarily true because mini-batch gradient descent can search through weight space due to noise and potentially escape bad local minima.

---

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



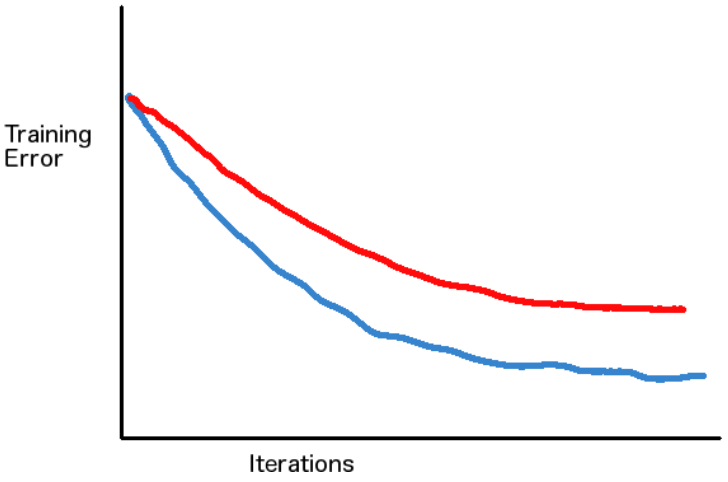
She was not sure if this was the best she could do. So she tried a **smaller** learning rate. Which of the following error curves (shown in red) might she observe now? Select the two most likely plots.

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

Lecture 6 Quiz

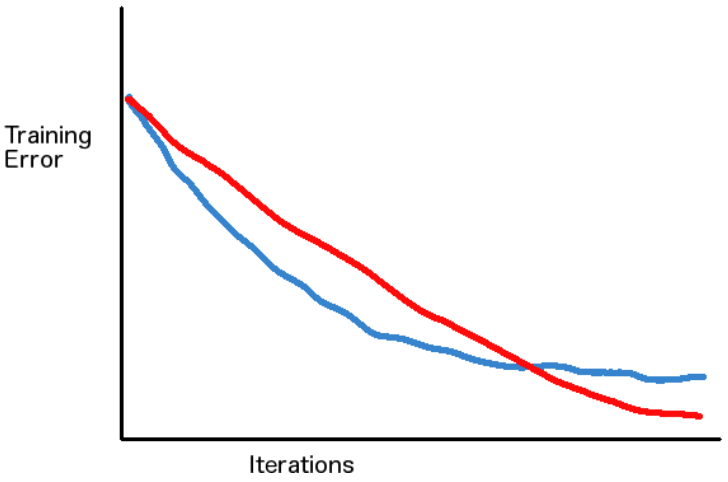
Quiz, 5 questions

5/5 points (100%)



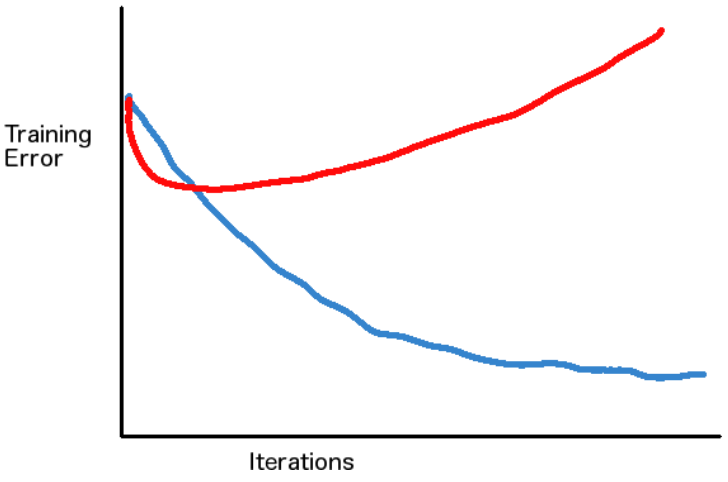
Correct

A smaller learning rate may lead to slower convergence.



Correct

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

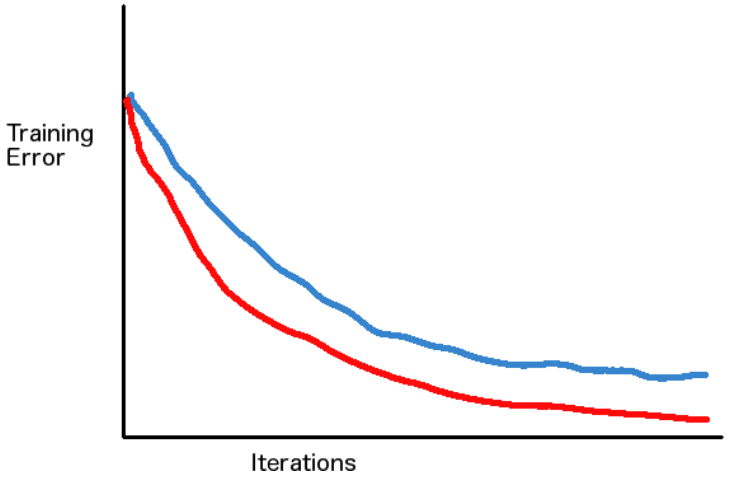


Un-selected is correct

Lecture 6 Quiz

5/5 points (100%)

Quiz, 5 questions



Un-selected is correct

---

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?

- [ ] 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

- [ ] 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