3.

a.

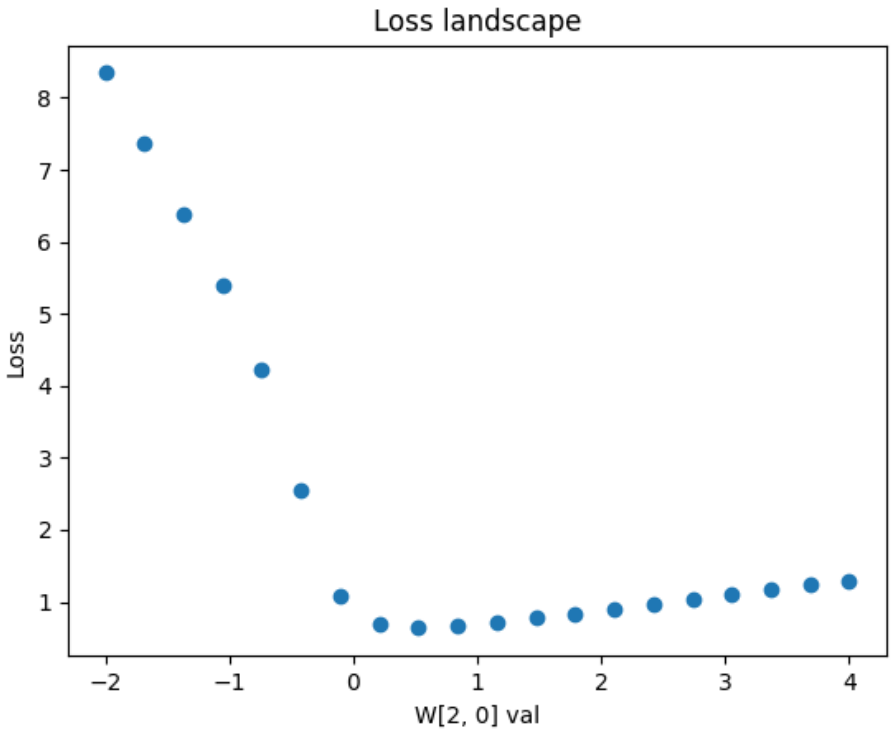


Figure 4: batch size 10

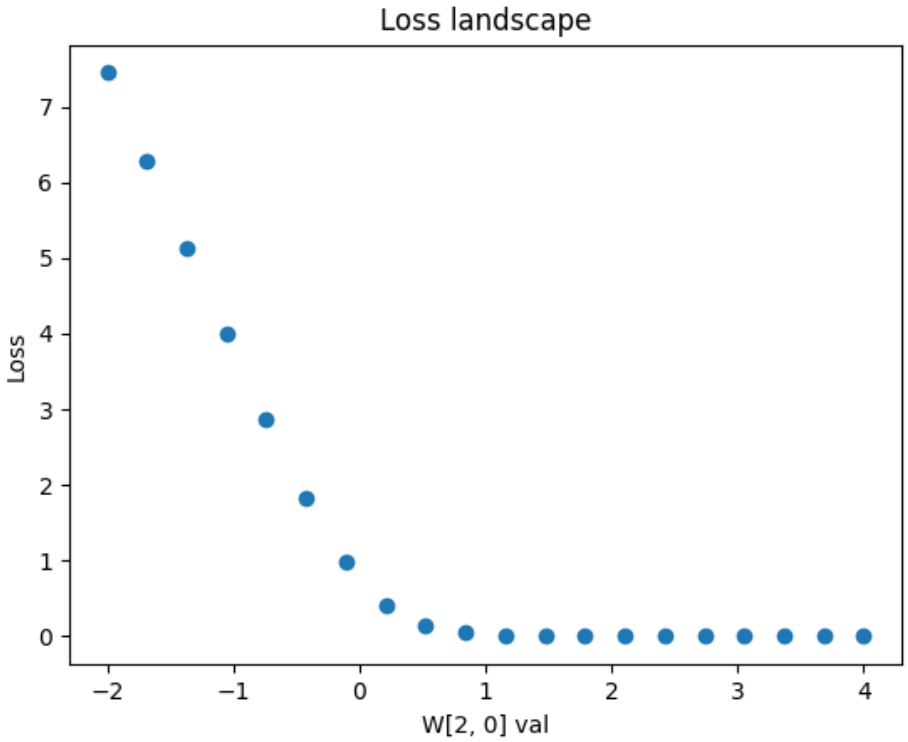


Figure 5: batch size 1

The batch size of 10 had an standard deviation of 0.16 meaning it was more accurate because it sampled from a bigger data. The batch size of 1 had a standard deviation of 2.16 and that is because the there is less data the information is collected from so the accuracy is less.

b. Batch gradient descent is when you use all of the data and this is the most accurate representation which is why the standard deviation is so low, but did take the longest to calculate. Stochastic gradient descent did a lot of fast calculations, but the standard deviation was very large. Minibatch gradient decent is in between batch gradient descent and stochastic gradient descent having to trade speed and accuracy.