Assignment3-Part3

Mengtong Zhang
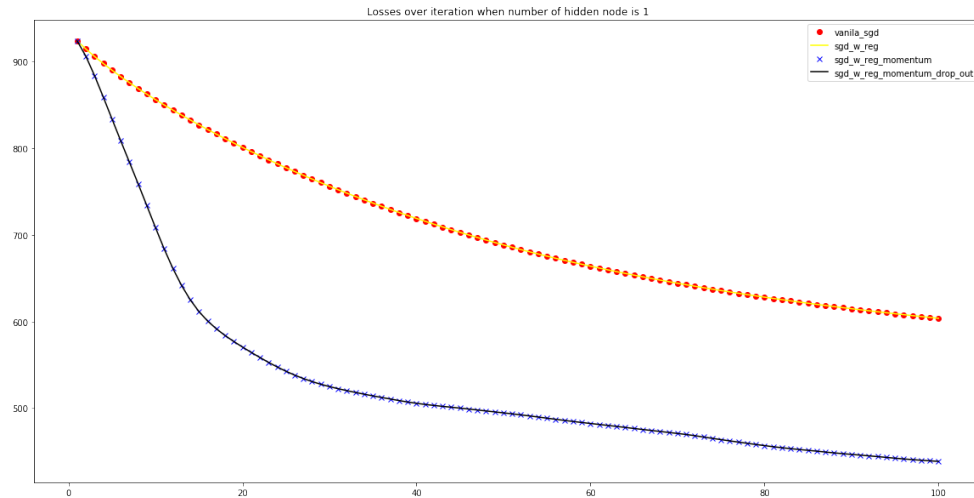
April 2020

Collaborate with Shaoyu Feng and Yunjia Zeng
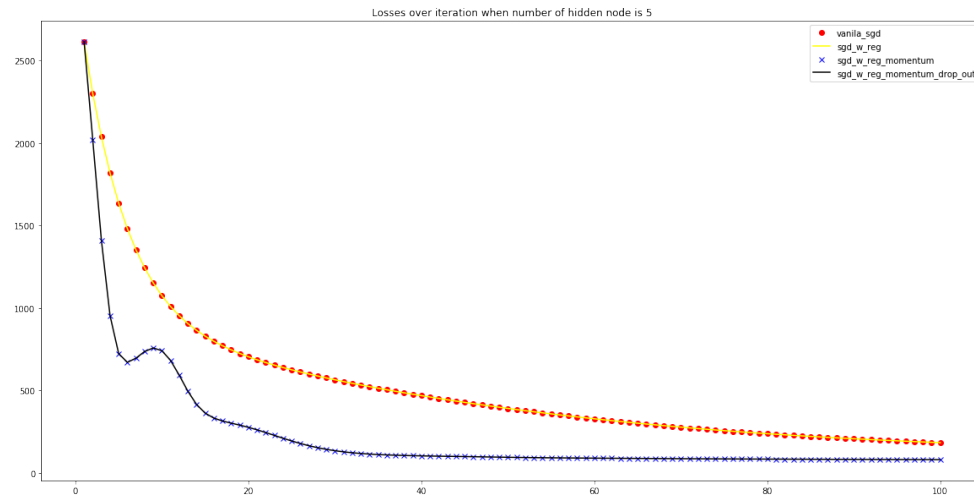
1 Question 5

1.1 Part1:ReLu

A.

a. SGD with momentum and dropout/SGD with momentum has the best performance.
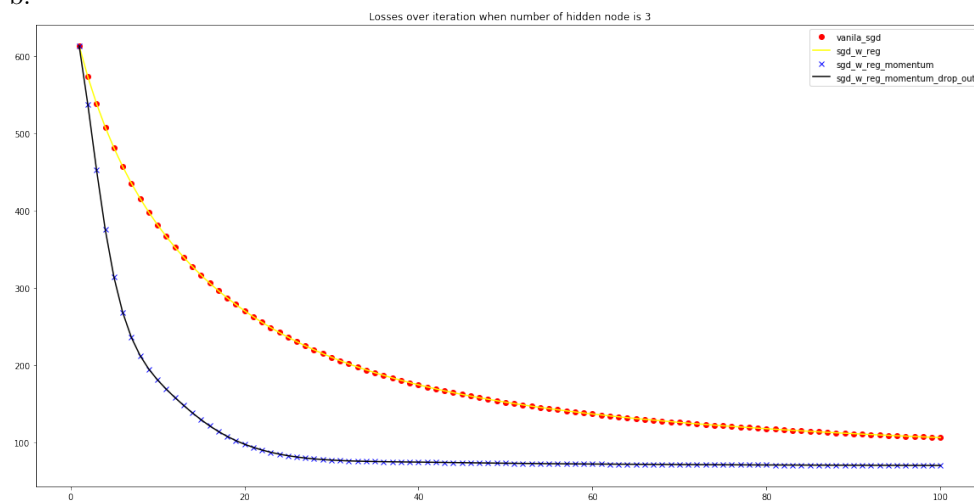
b. Because it prevent convergence to local minimum by using momentum and prevent overfitting by dropout. So the losses drop really fast and converge very quickly.

   B.

Losses over iteration when number of hidden node is 5

a.SGD with momentum and dropout/SGD with momentum has the best performance.
b.

Losses over iteration when number of hidden node is 3

It seems node=3 has better performance. Perhaps it's because less node has prevented overfitting and fit the data better.

C.

Here is the time in seconds for each algorithm at different number of nodes.
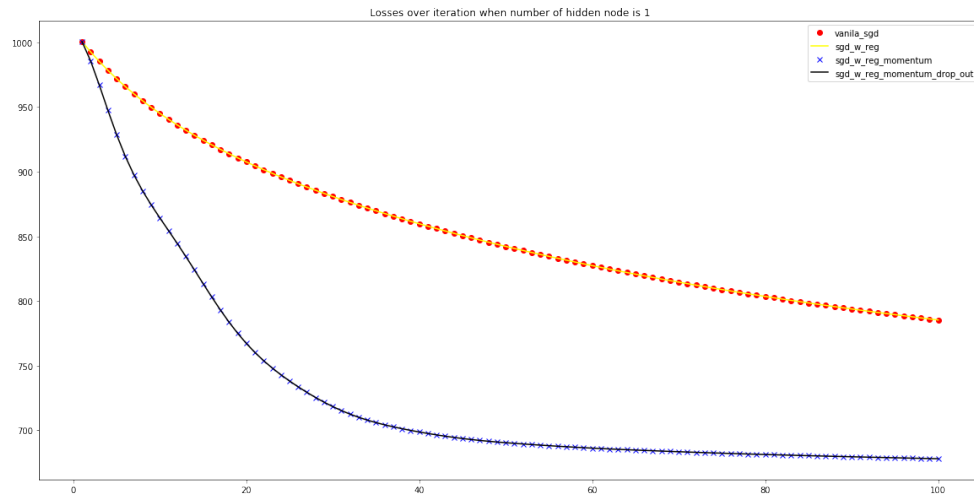
	node=1	node=3	node=5
Vanilla SGD	0.70	1.63	2.65
SGD with reg	0.64	1.57	2.27
SGD with reg momentum	0.73	1.70	2.26
SGD with reg momentum dropout	0.53	1.63	2.43

And this is the loss.

	node=1	node=3	node=5
Vanilla SGD	634	120	251
SGD with reg	634	120	251
SGD with reg momentum	462	70	83
SGD with reg momentum dropout	462	70	83

1.2   Part2:Sigmoid

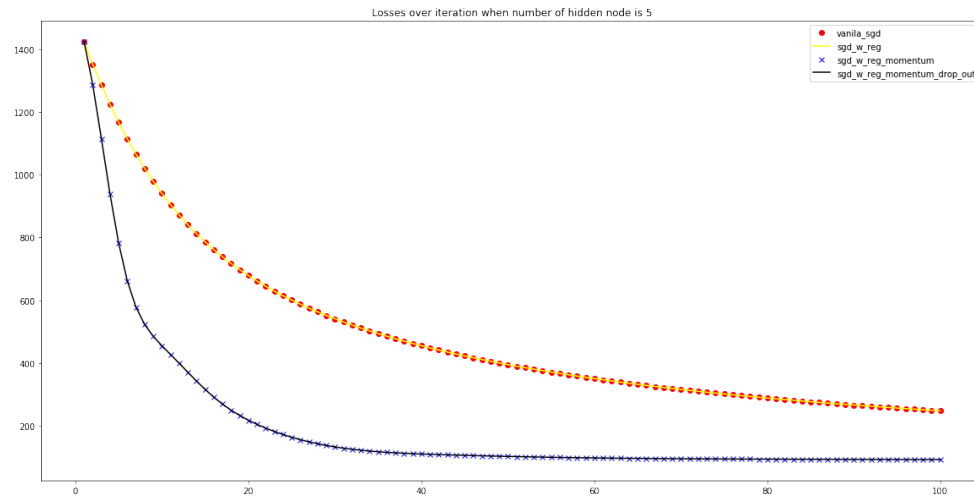A.

Losses over iteration when number of hidden node is 1

a. SGD with momentum and dropout/SGD with momentum has the best performance.
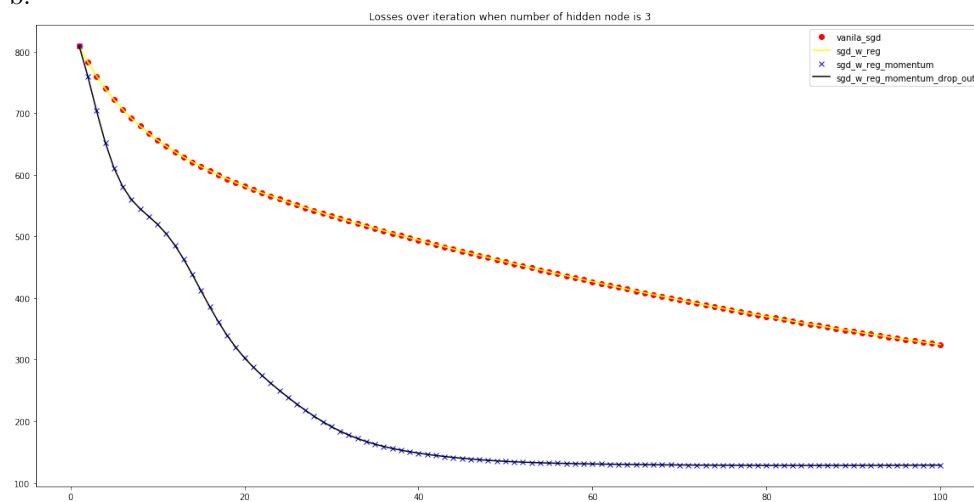
b. Because it prevent convergence to local minimum by using momentum and prevent overfitting by dropout. So the losses drop really fast and converge very quickly.

B.

Losses over iteration when number of hidden node is 5

a.SGD with momentum and dropout/SGD with momentum has the best performance.

b.

Losses over iteration when number of hidden node is 3

It seems node=5 has better performance. Perhaps it's because 5 nodes make the model more flexible to fit the data better.

C.

	node=1	node=3	node=5
Vanilla SGD	0.07	0.09	0.1
SGD with reg	0.06	0.07	0.08
SGD with reg momentum	0.06	0.08	0.1
SGD with reg momentum dropout	0.08	0.08	0.08

D.

4

Here we make a comparison between sigmoid and ReLu.The data of ReLU are in parenthesis.

Time:

|	node=1	node=3	node=5
Vanilla SGD	0.70(0.07)	1.63(0.09)	2.65(0.1)
SGD with reg	0.64(0.06)	1.57(0.07)	2.27(0.08)
SGD with reg momentum	0.73(0.06)	1.70(0.08)	2.26(0.1)
SGD with reg momentum dropout	0.53(0.08)	1.63(0.08)	2.43(0.08)

Loss:

	node=1	node=3	node=5
Vanilla SGD	634(807)	120(380)	251(298)
SGD with reg	634(807)	120(380)	251(298)
SGD with reg momentum	462(682)	70(128)	83(93)
SGD with reg momentum dropout	462(682)	70(128)	83(93)

2 Question 6

2.1 1.a

For one node and one hidden layer, here are all the possibilities:
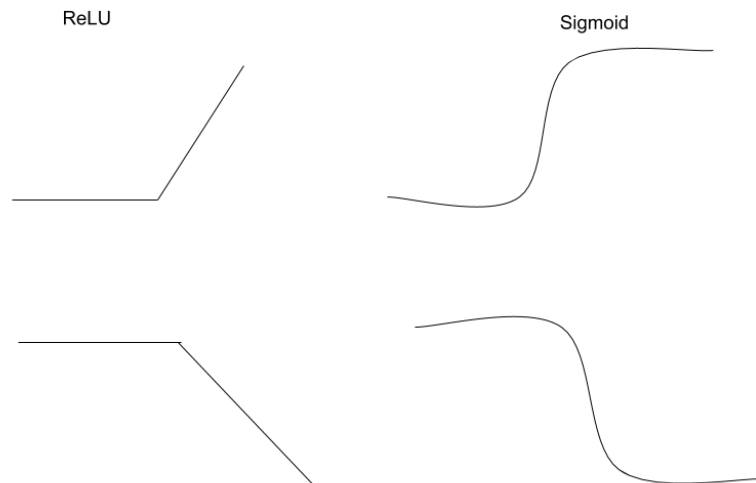
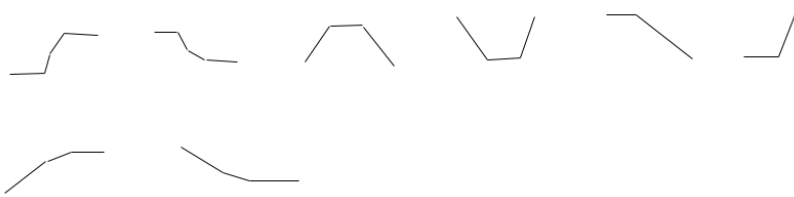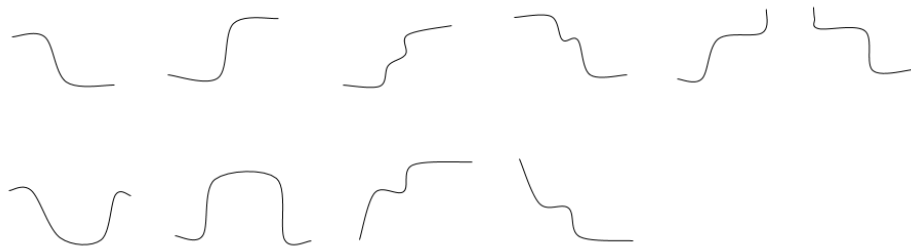ReLU   Sigmoid

2.2   1.b

After tried by coding different probabilities, I got states for different signs of
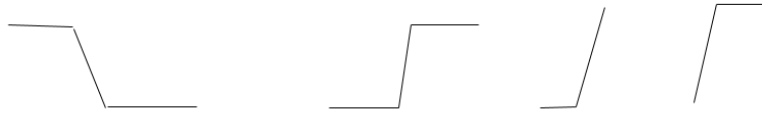weights as below:

ReLU

Sigmoid

2.3 2.a

ReLU

Sigmoid

2.4   2.b

ReLU

Sigmoid