

H3P3

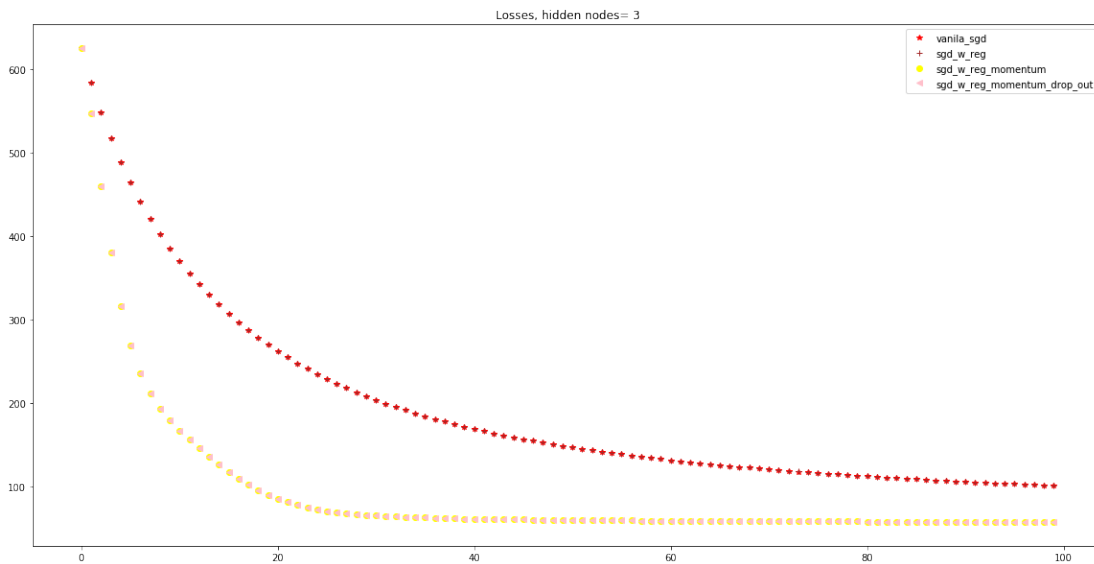
Guanzhi Wang

4/15/2020"

Collaborate with Shaoyu Feng

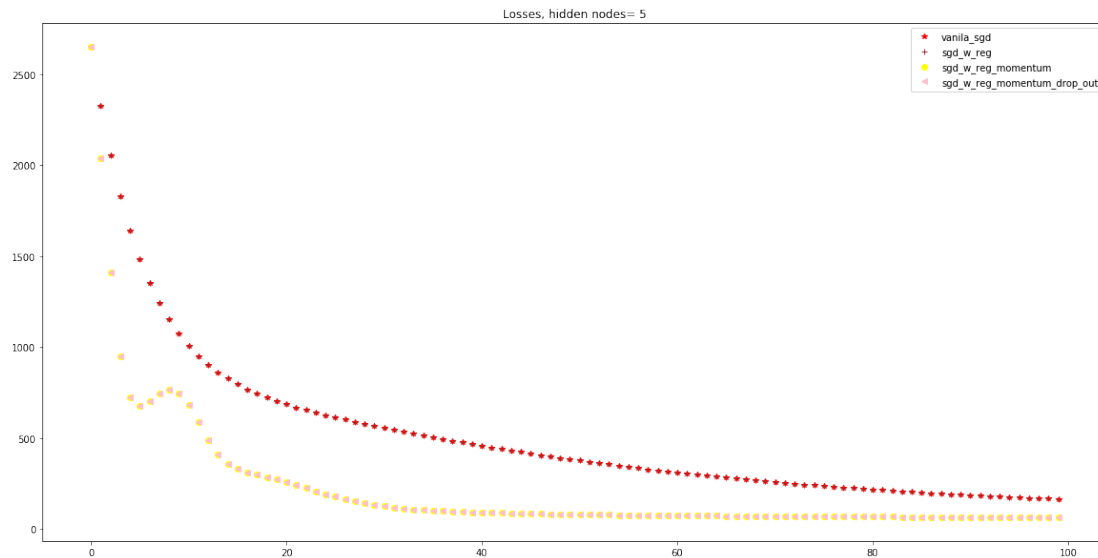
Question 6

Part1. A



- (a) From the graph we can see the sgd_w_reg_momentum and sgd_w_reg_momentum_drop_out have the almost the same performance and both the best.
- (b) Because by using the momentum, the gradients are speeded up in the direction of converge when processing. Thus, it could drop the loss faster.

Part1. B



- (a) Similar to A, sgd_w_reg_momentum and sgd_w_reg_momentum_drop_out still performance the same and still better than first two methods.
- (b) No. Since hidden nodes=3 is already enough. When there are 5 nodes, overfitting occurs.

(c)

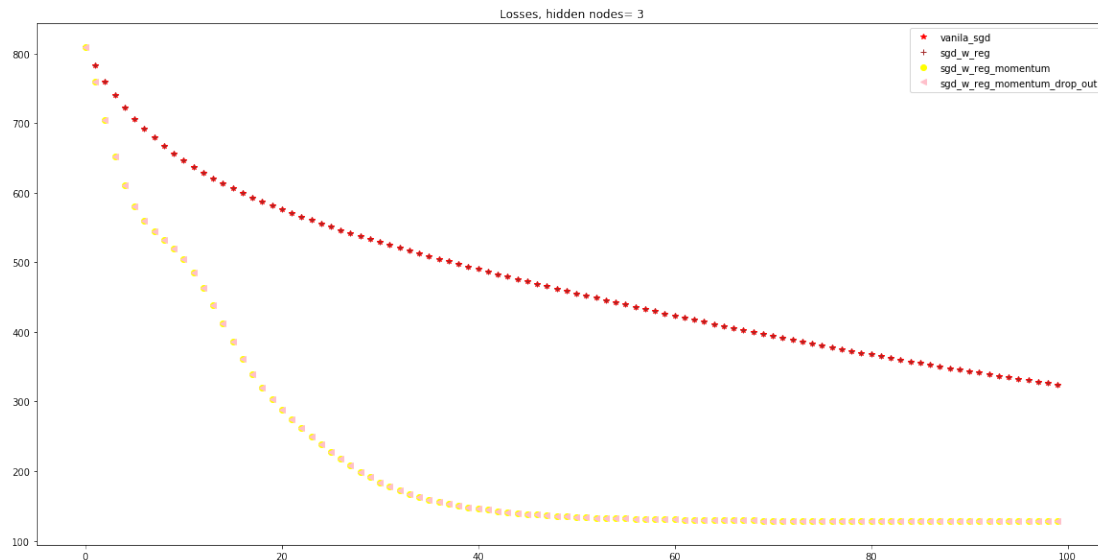
```

=====Vanilla_SGD=====
0.9085500000000053 seconds
=====sgd_w_reg=====
0.8882940000000019 seconds
=====sgd_w_reg_momentum=====
=====
0.9254169999999959 seconds
=====sgd_w_reg_momentum_drop_out=====
0.8651150000000003 seconds

```

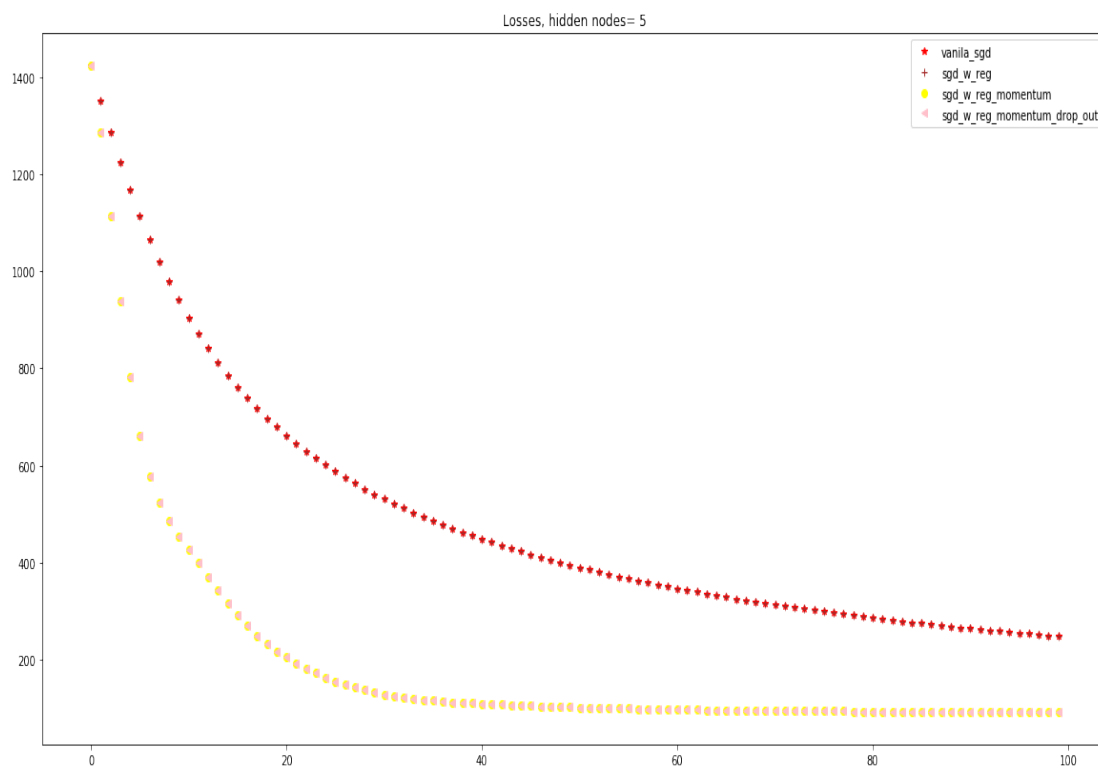
Part 2

(a)



Same as part 1, sgd_w_reg_momentum and sgd_w_reg_momentum_drop_out still performance the same and still better than first two methods.

(b)



Nodes=5 are better than nodes=3. Which means complexity helps model to learn.

(c)

=====
 =====Vanilla_SGD=====

0.9055129999999991 seconds

```

=====sgd_w_reg=====
0.8649379999999987 seconds
=====sgd_w_reg_momentum=====
0.872522 seconds
=====sgd_w_reg_momentum_drop_out=====
0.8725799999999992 seconds

```

(d)

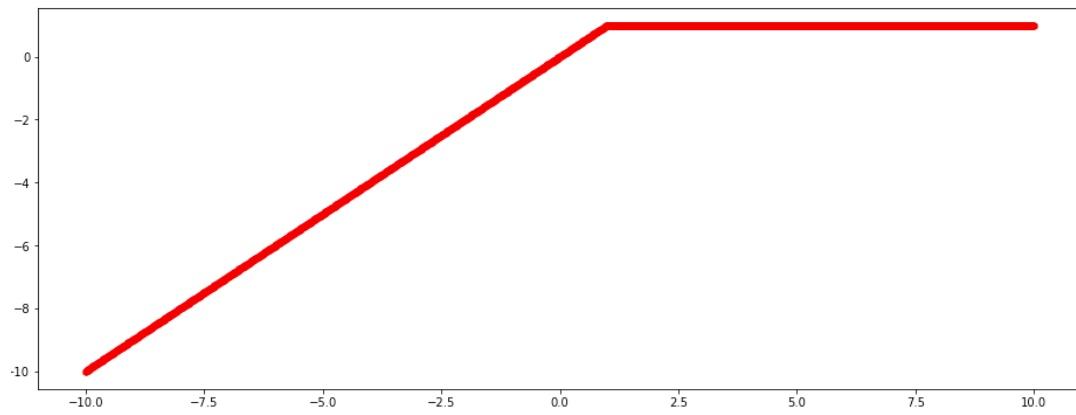
Active func	Model	Loss	Runtime
ReLU	Vanilla_SGD n =3	116.03	0.90
	sgd_w_reg n =3	116.04	0.88
	sgd_w_reg_momentum n =3	58.23	0.92
	sgd_w_reg_momentum_drop_out n =3	58.23	0.86
	Vanilla_SGD n =5	236.92	1.46
	sgd_w_reg n =5	236.93	1.33
	sgd_w_reg_momentum n =5	69.58	1.42
	sgd_w_reg_momentum_drop_out n =5	69.58	1.36
Sigmoid	Vanilla_SGD n =3	382.72	0.86
	sgd_w_reg n =3	382.75	0.79
	sgd_w_reg_momentum n =3	127.63	0.80
	sgd_w_reg_momentum_drop_out n =3	127.63	0.79
	Vanilla_SGD n =5	293.19	1.34
	sgd_w_reg n =5	293.22	1.31
	sgd_w_reg_momentum n =5	89.78	1.33
	sgd_w_reg_momentum_drop_out n =5	89.78	1.30

Question 6

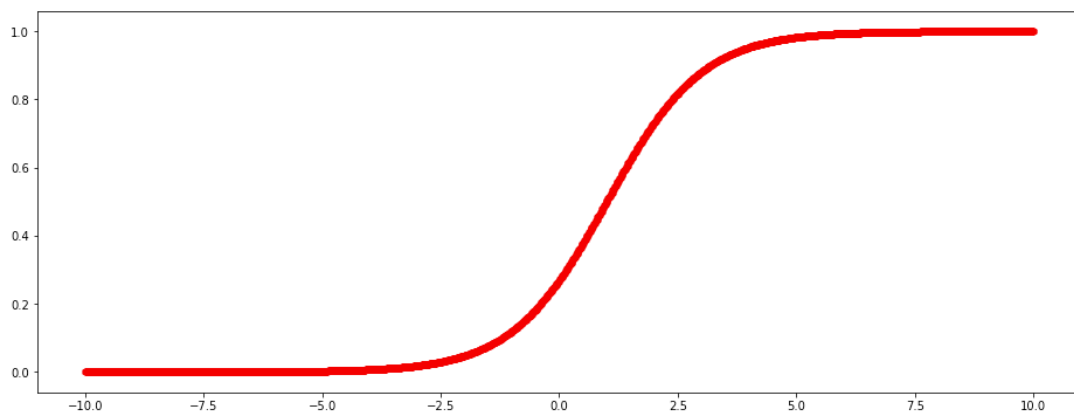
Please refer to the folders with pictures. Only brief screenshots are shown in the write up.

L=1, N=1:

ReLU

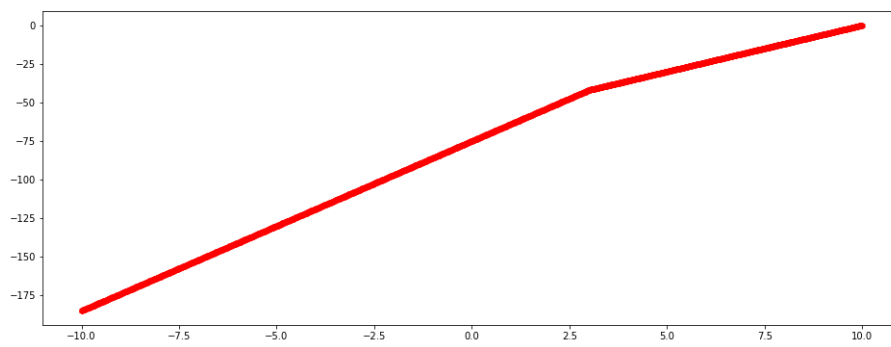


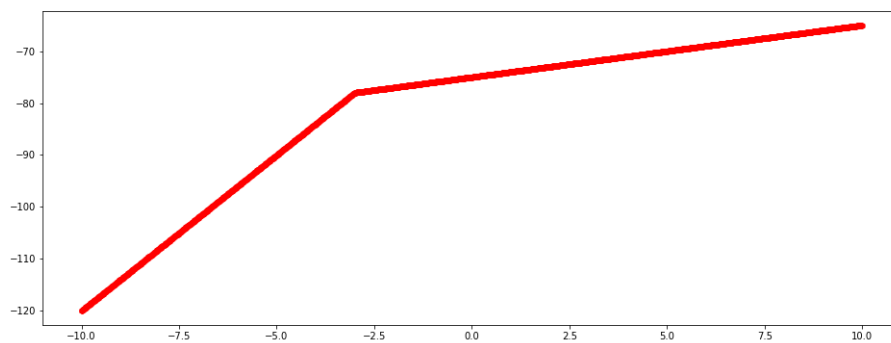
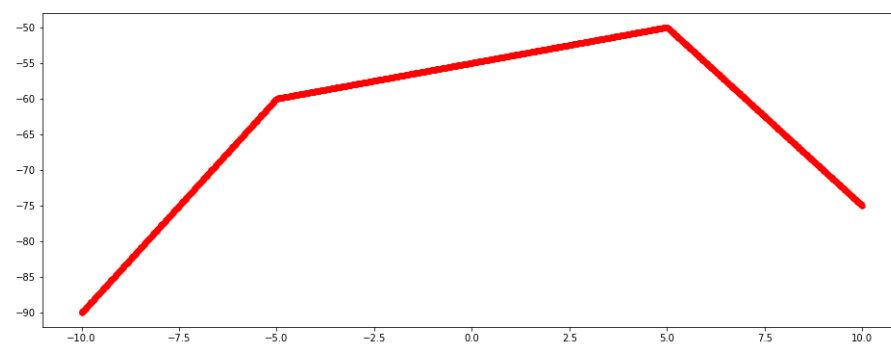
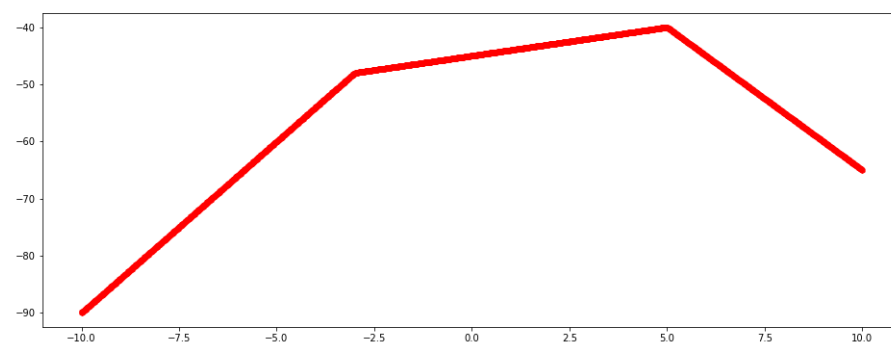
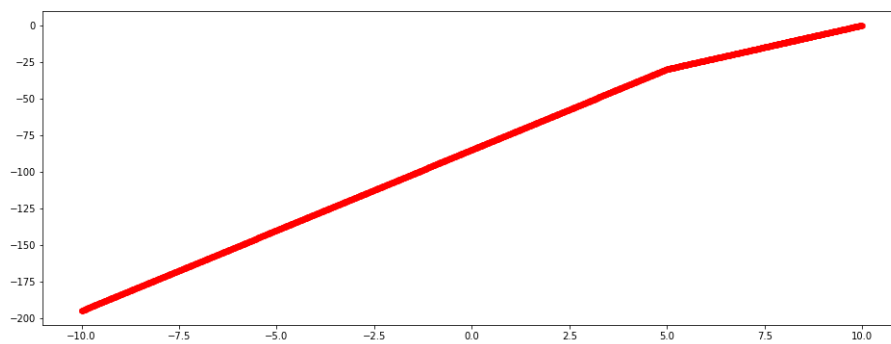
Sigmoid

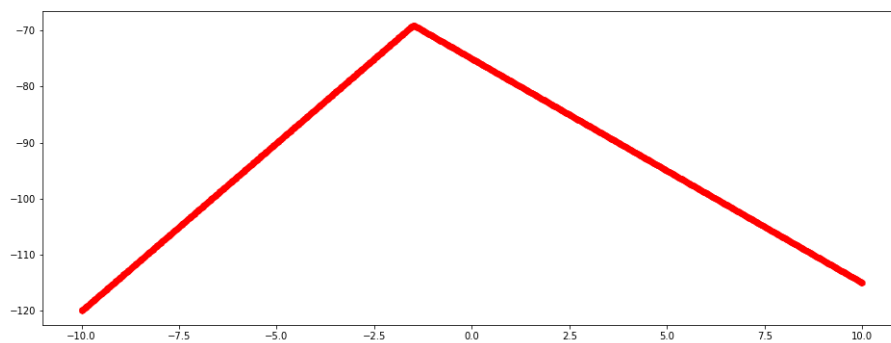
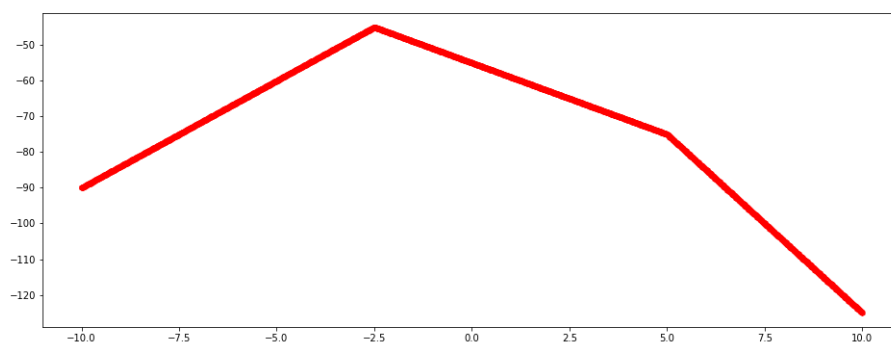
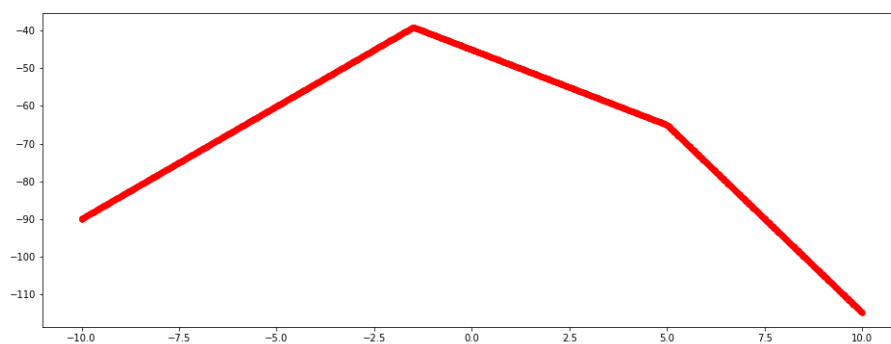
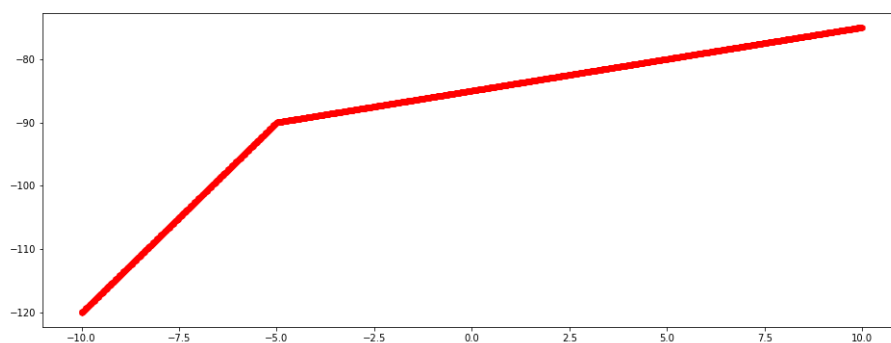


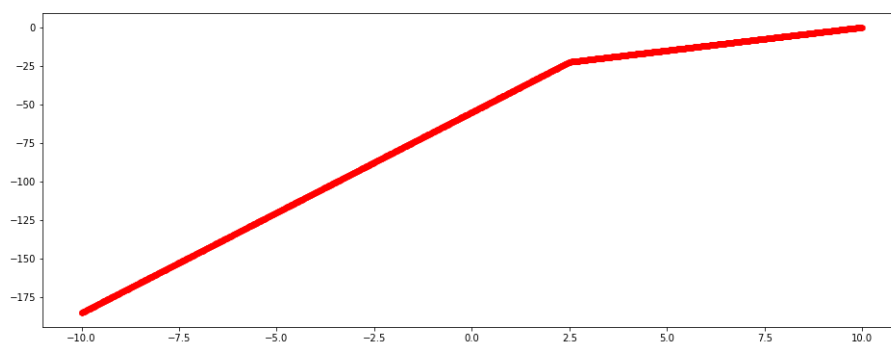
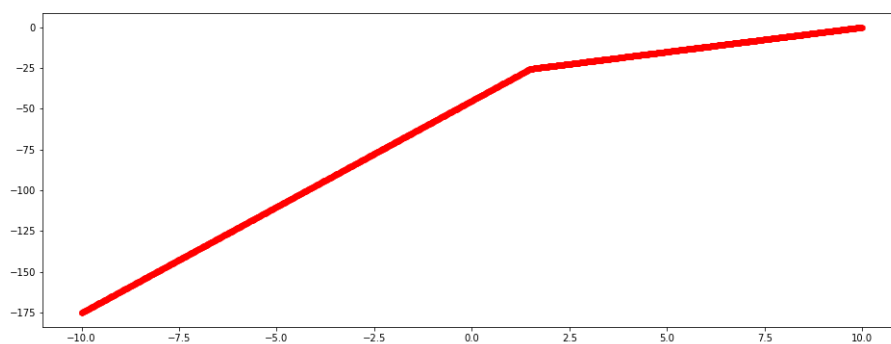
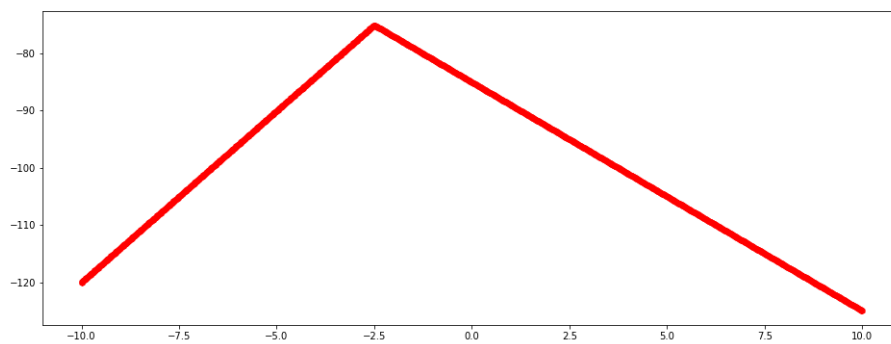
L=1, N=2:

ReLU

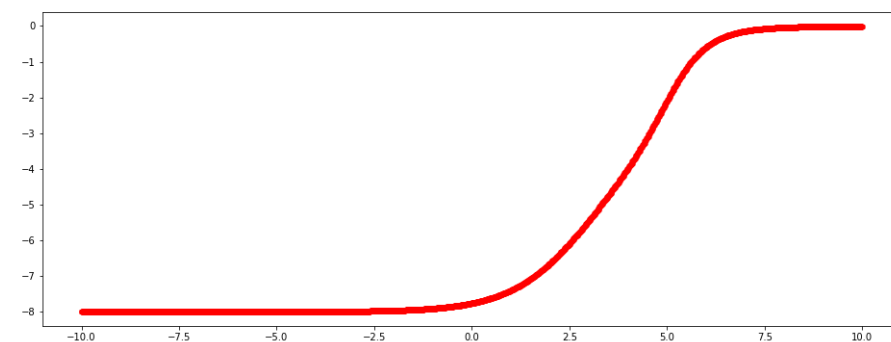


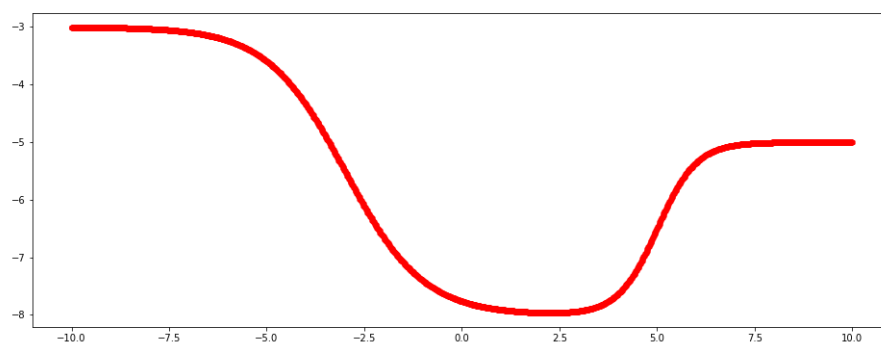
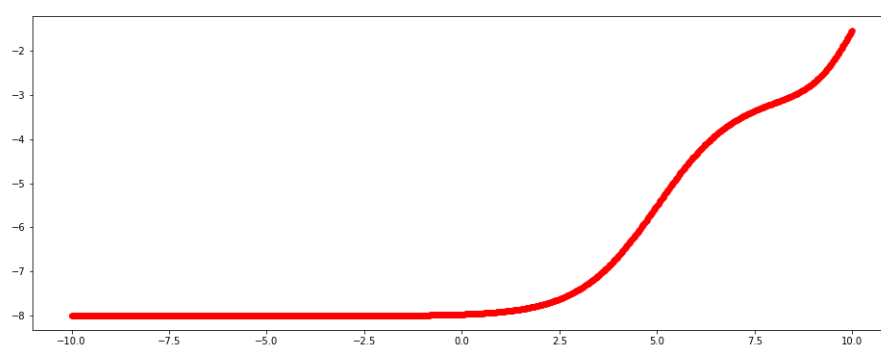
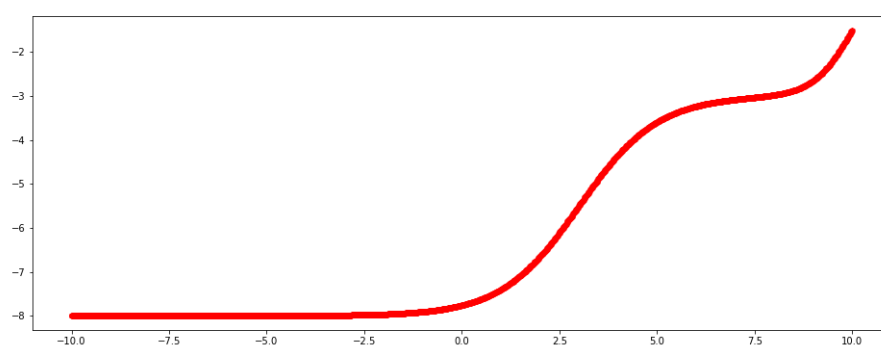
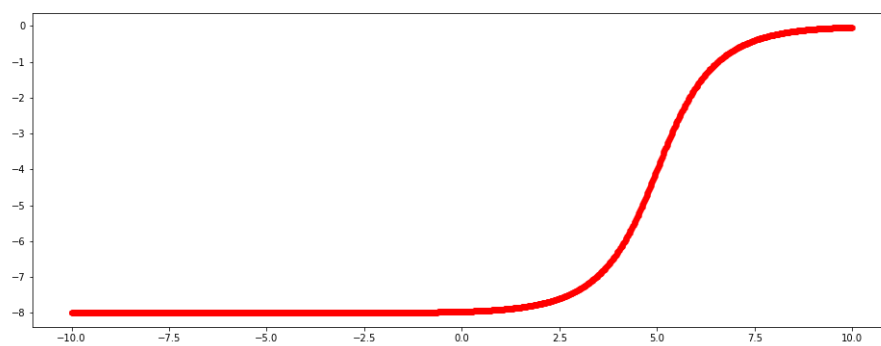


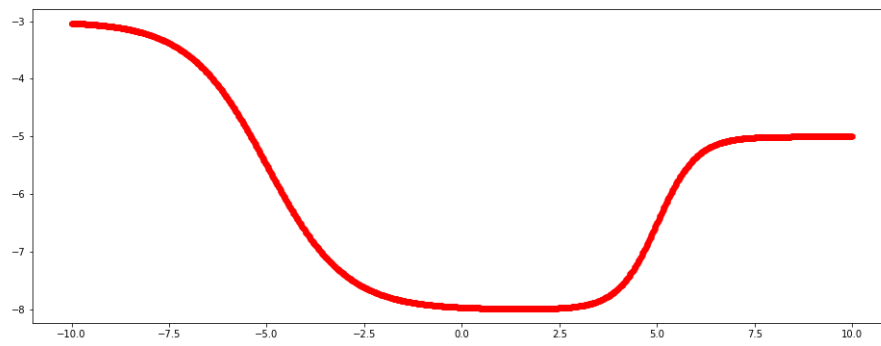




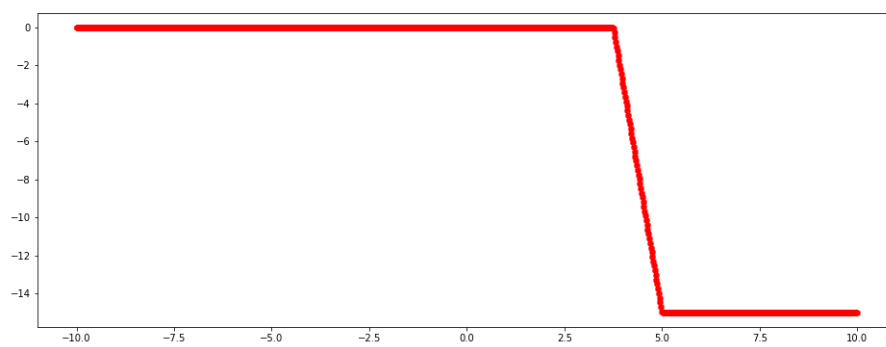
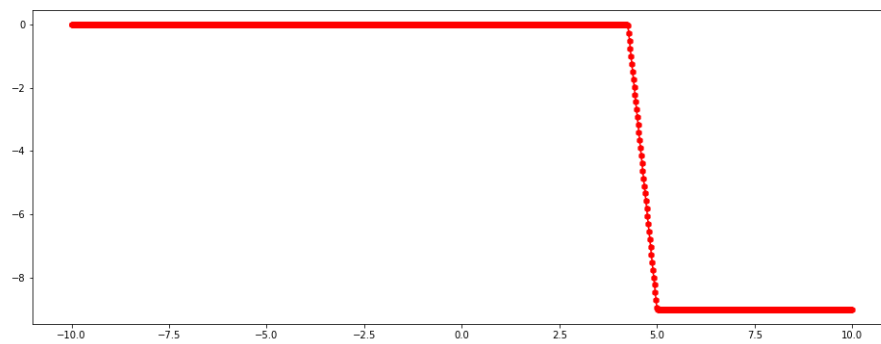
Sigmoid

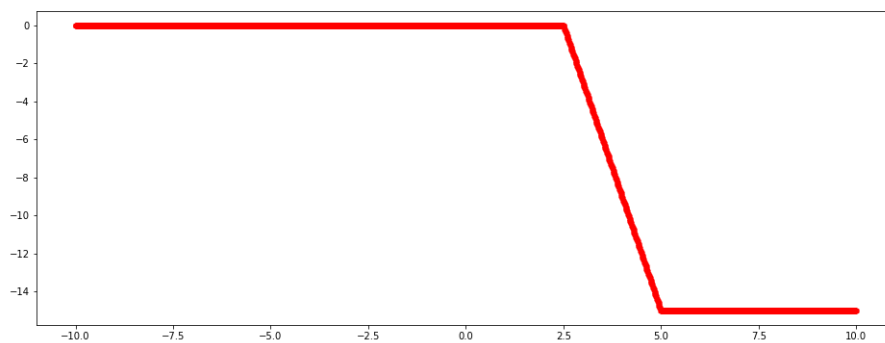
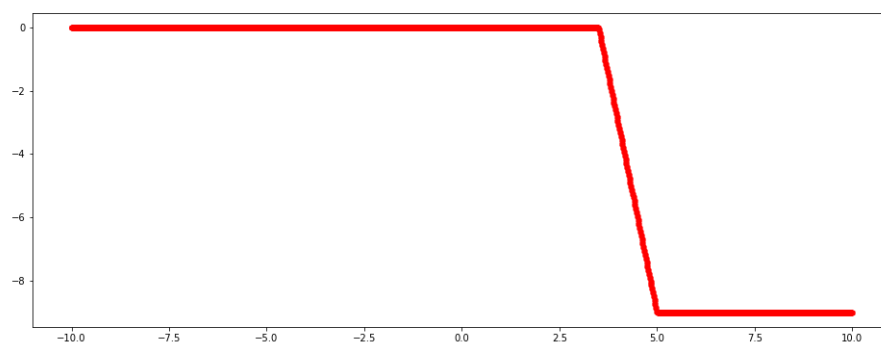
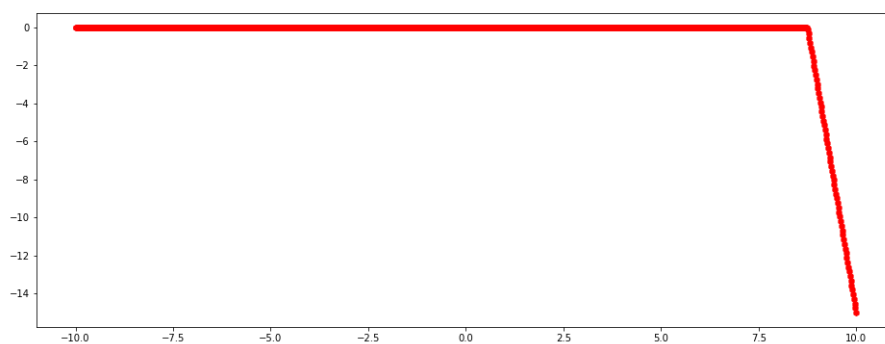
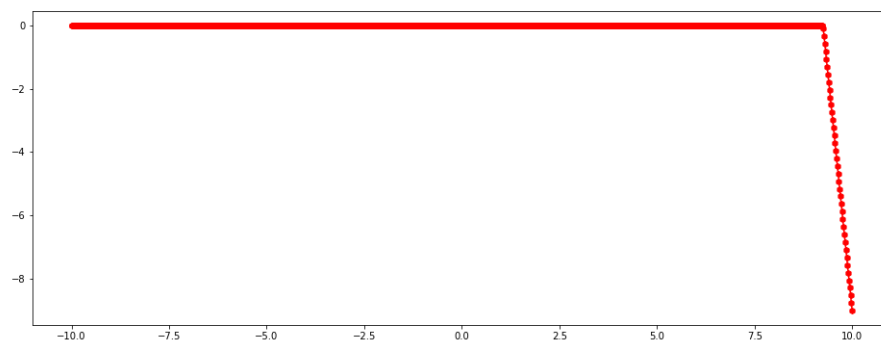




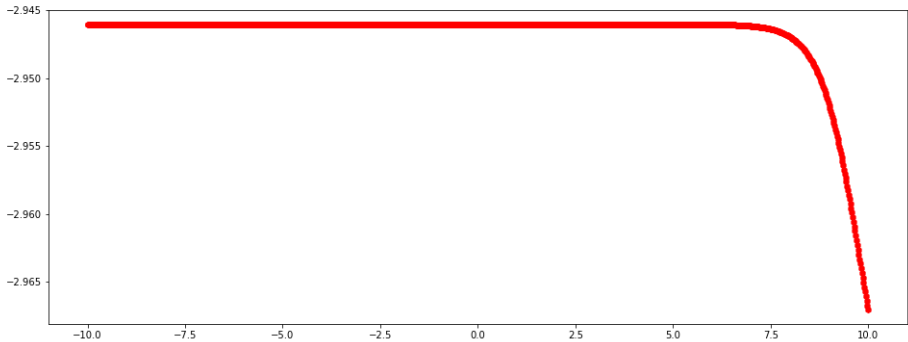
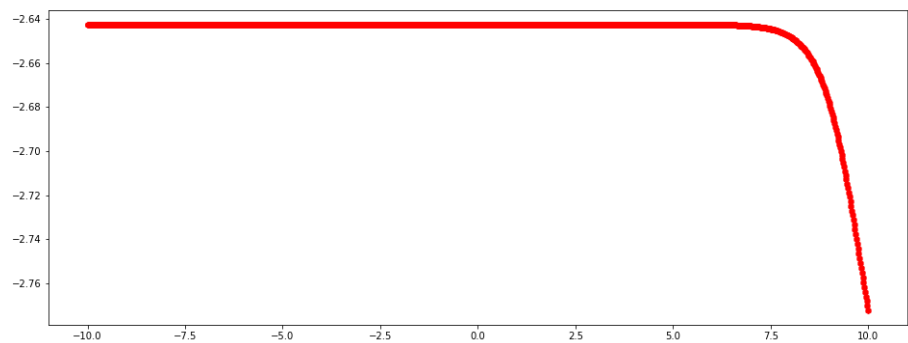
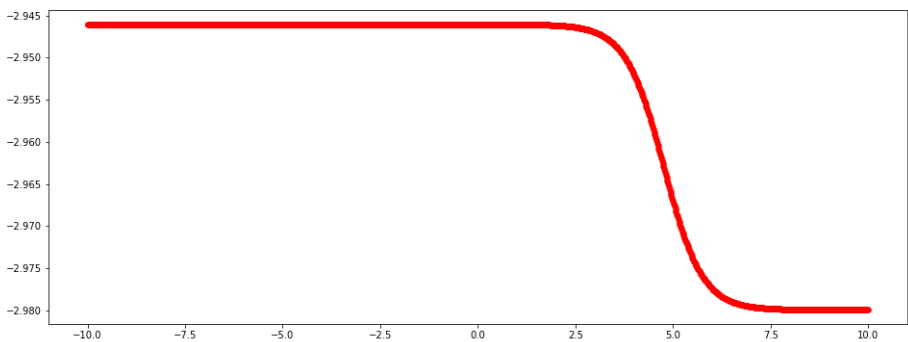
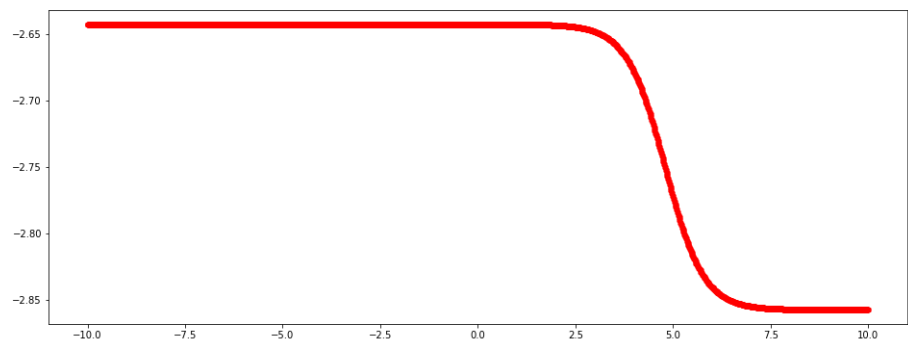


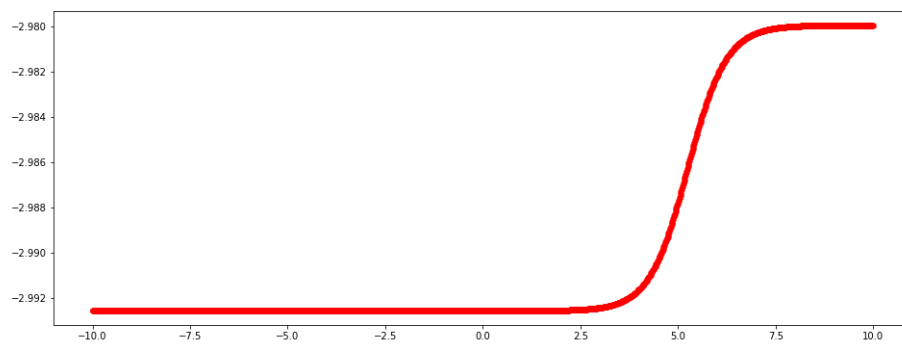
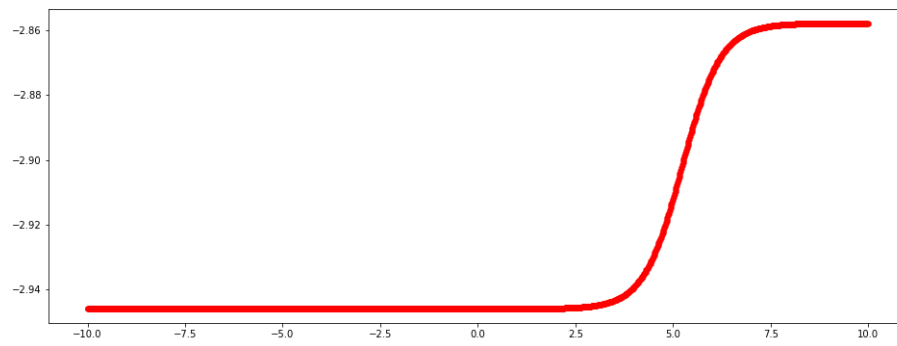
L=2, N=1:
ReLU



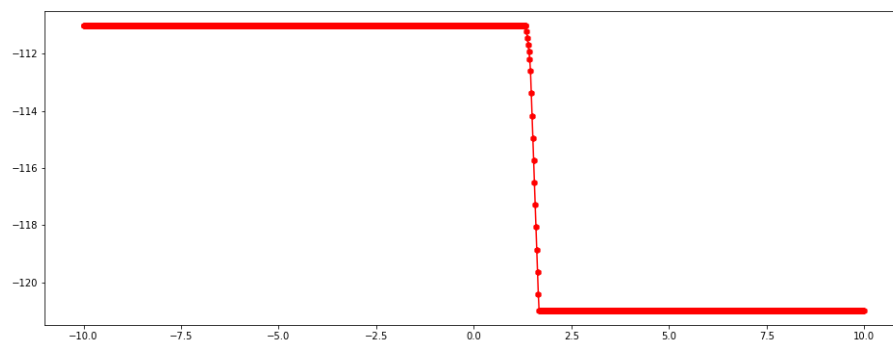


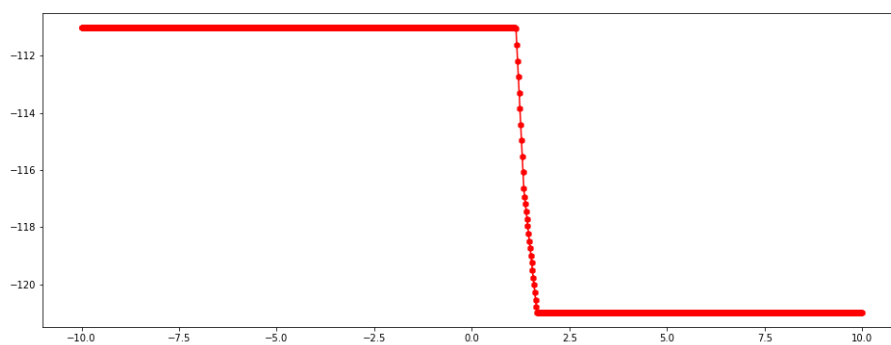
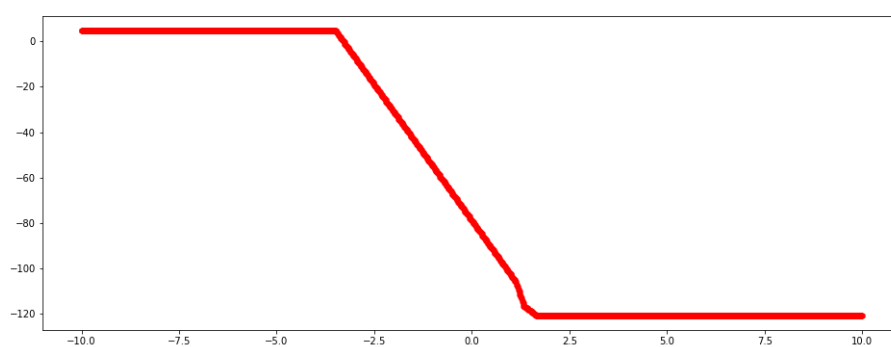
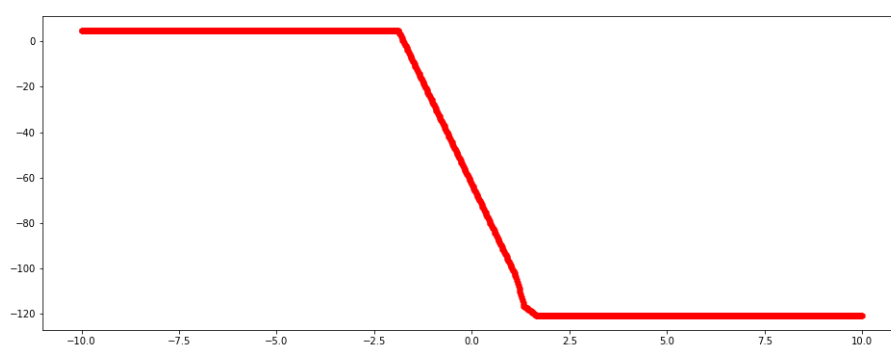
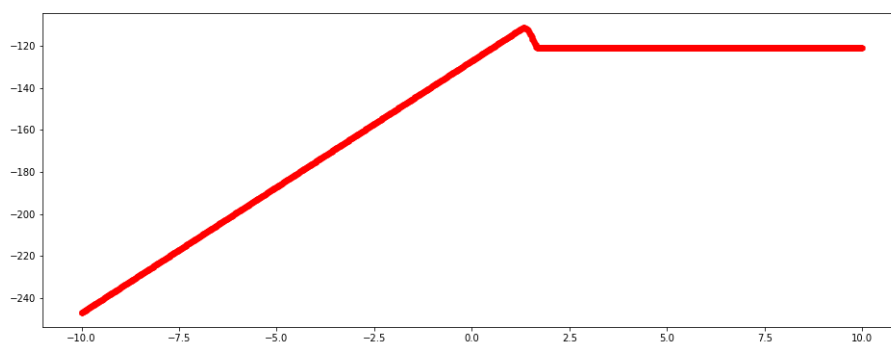
Sigmoid

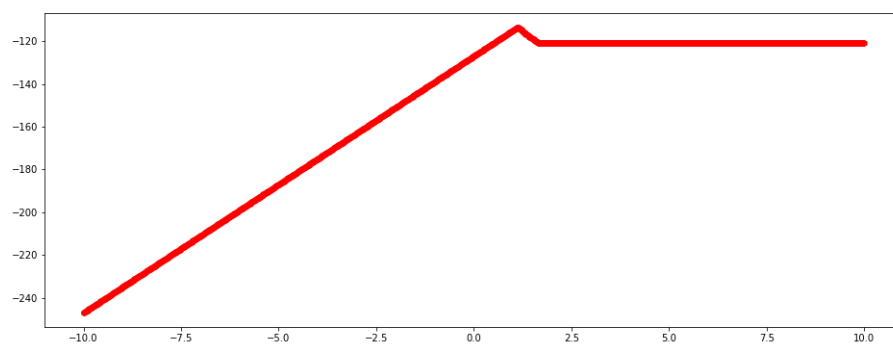




L=2, N=2:
ReLU







Sigmoid

