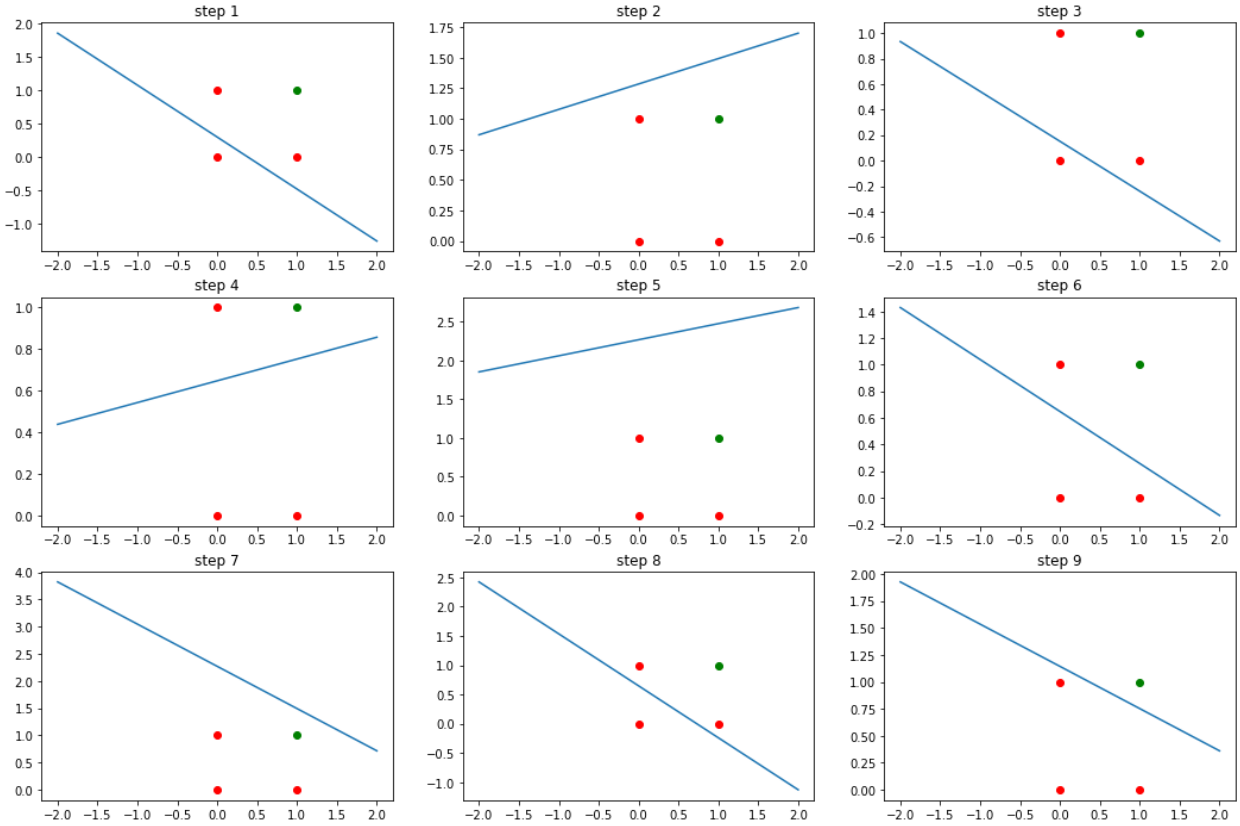
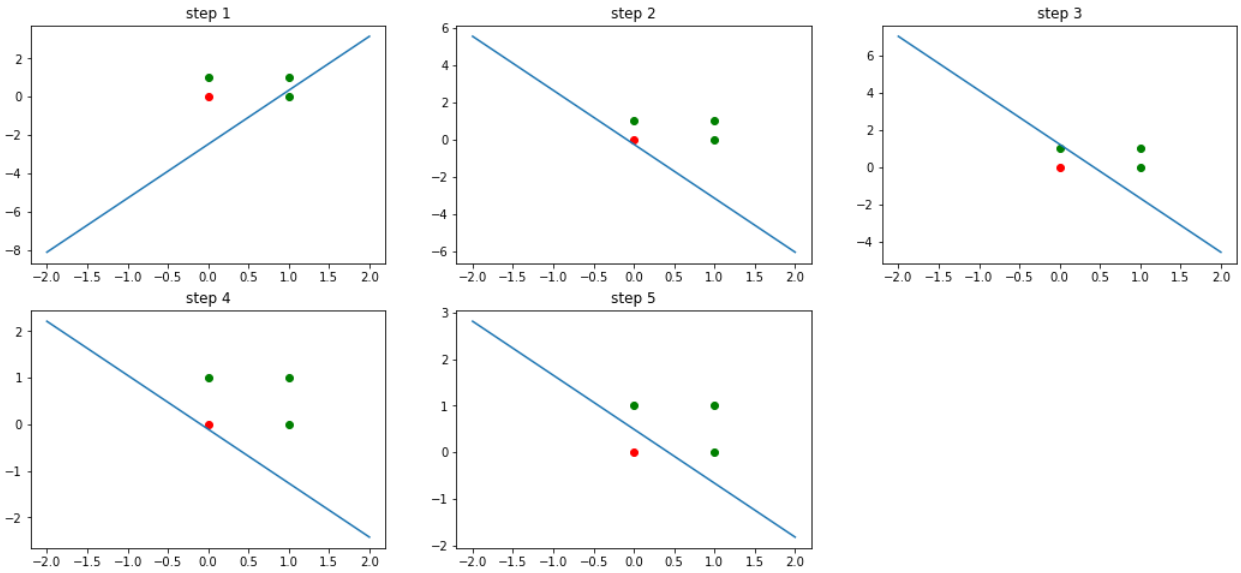
Part 1

Decision boundary of AND using PTA



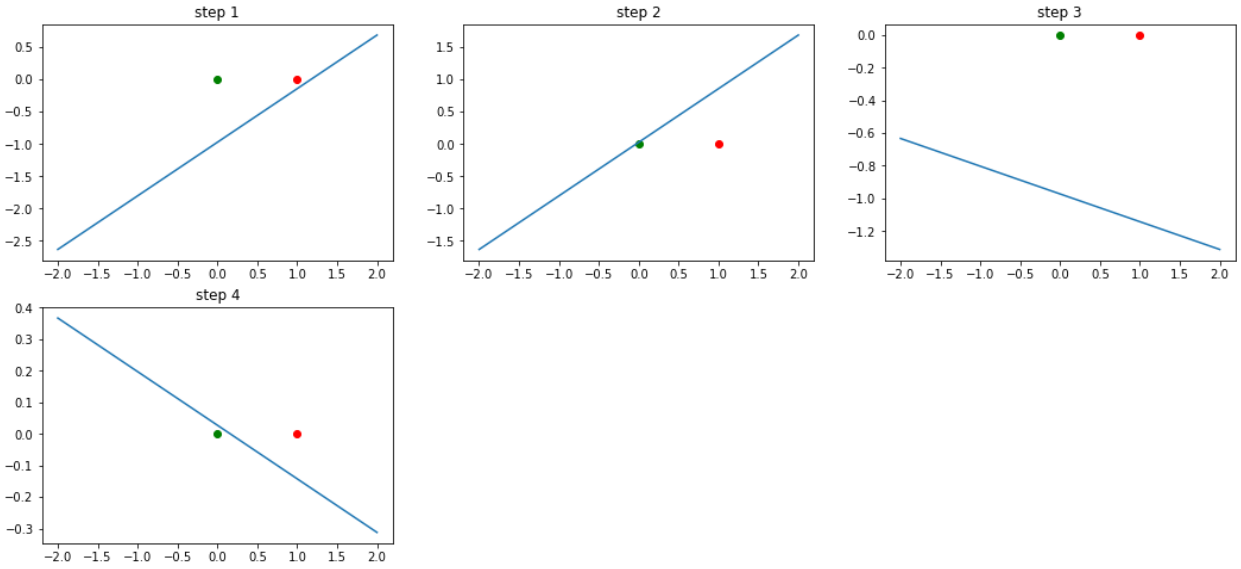
AND will converge after 9 steps using PTA.

Decision boundary of OR using PTA



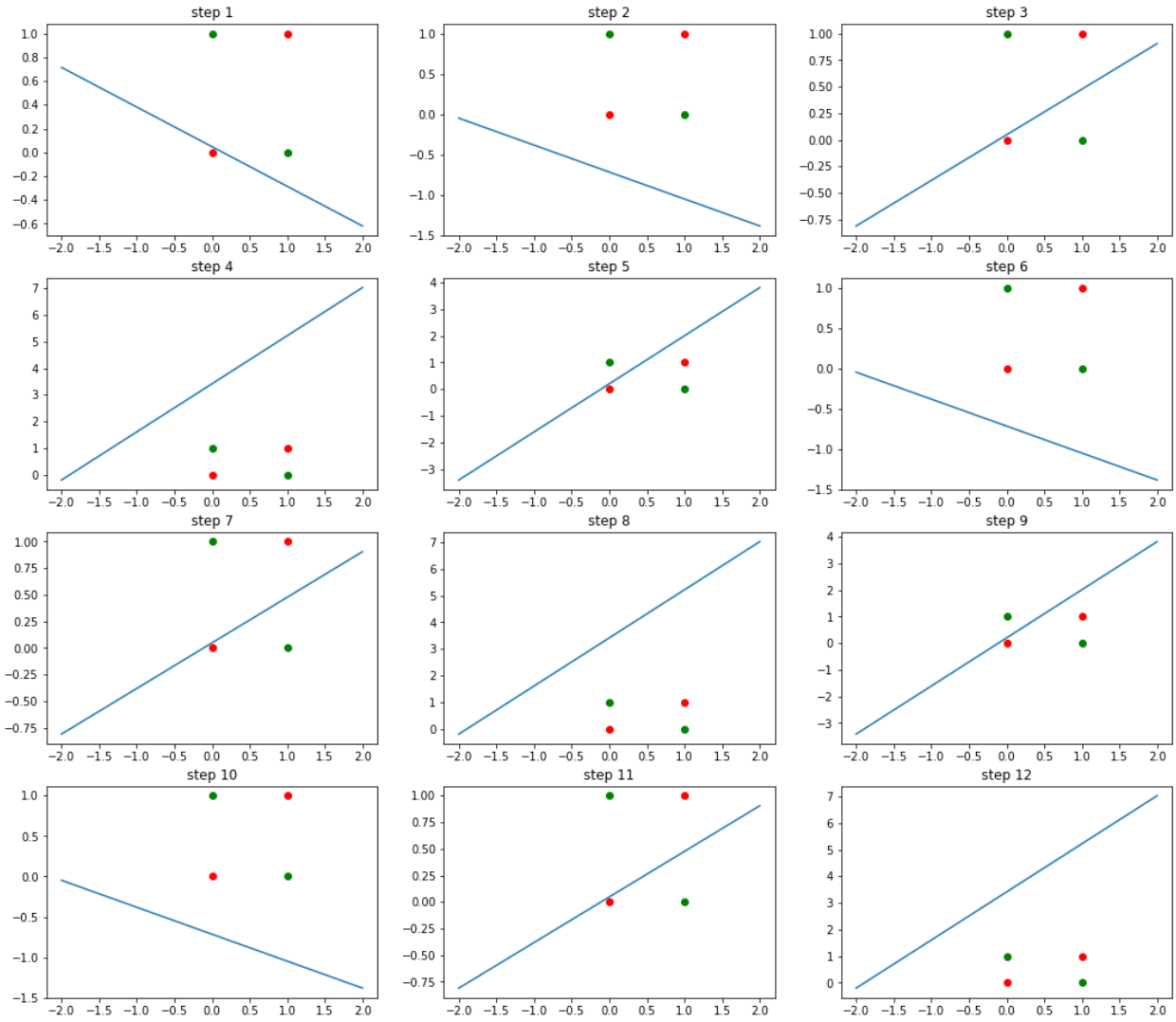
OR will converge after 5 steps.

Decision boundary of NOT using PTA



NOT will converge after 4 steps using PTA.

Decision boundary of XOR using PTA



As we can see XOR is oscillating around datapoints and if one points PTA correctly classified then other points misclassied.

its repeats decision boundary on datapoints after few iterations.

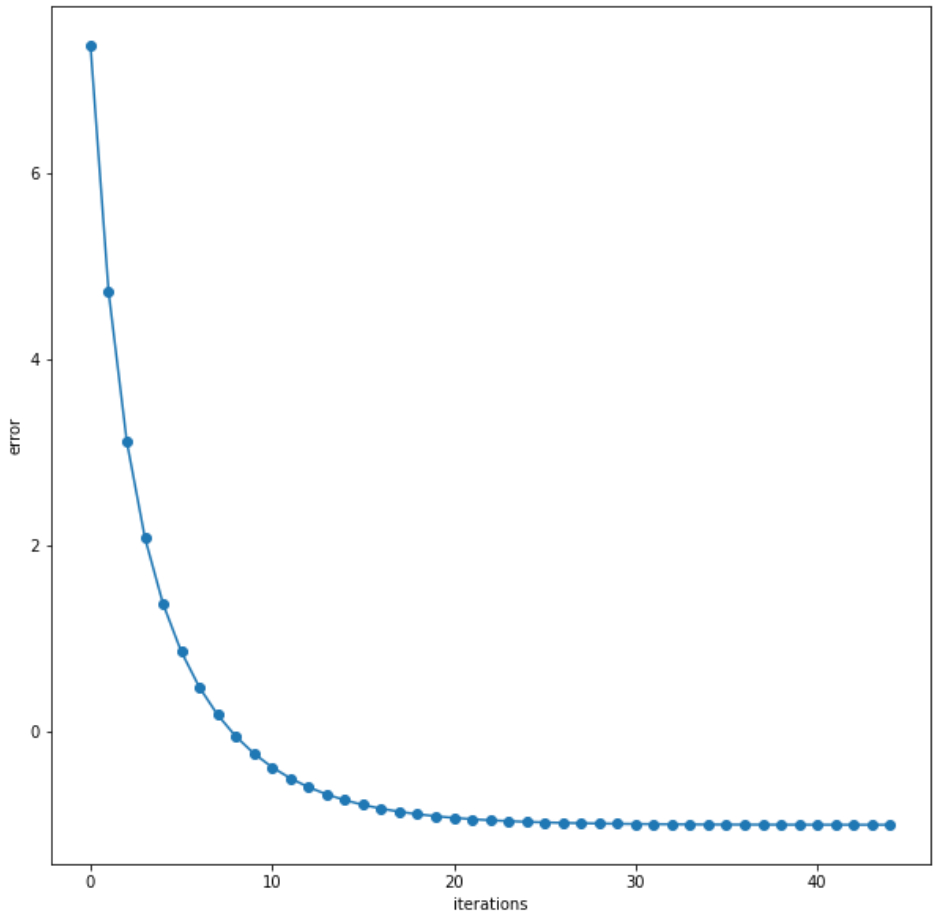
So we can see PTA will not converge on non linearly seperable dataset.

Part 2

```
#hyperparameters
learning_rate = 0.1
epochs = 1000
Y = 1
convergence_tolerance = 10**-4
```

2.c

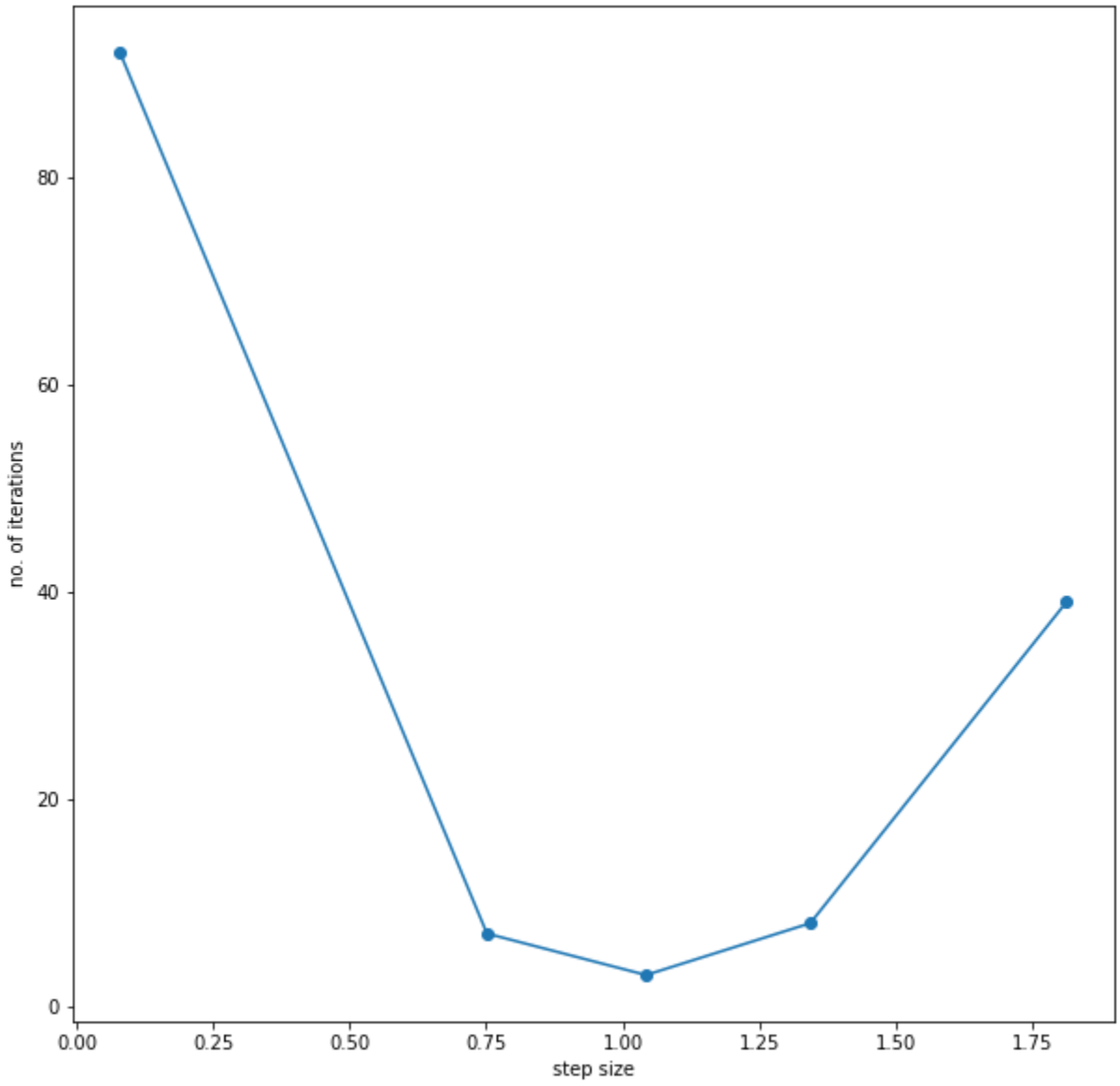
Iterations vs error plot



2.d

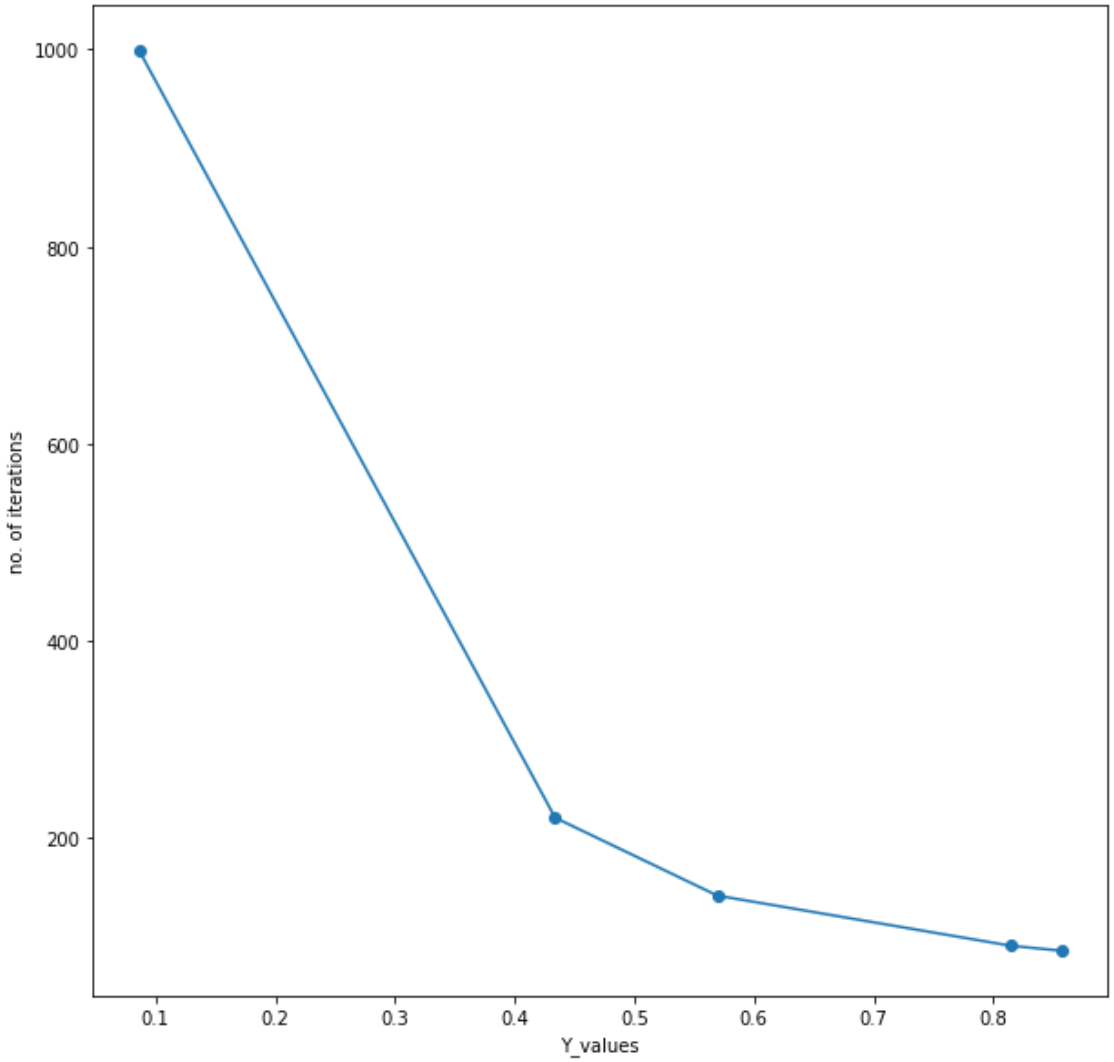
Step sizes list

```
array([0.0793962 , 0.75145342, 1.04280425, 1.34327902,
1.81265895])
```

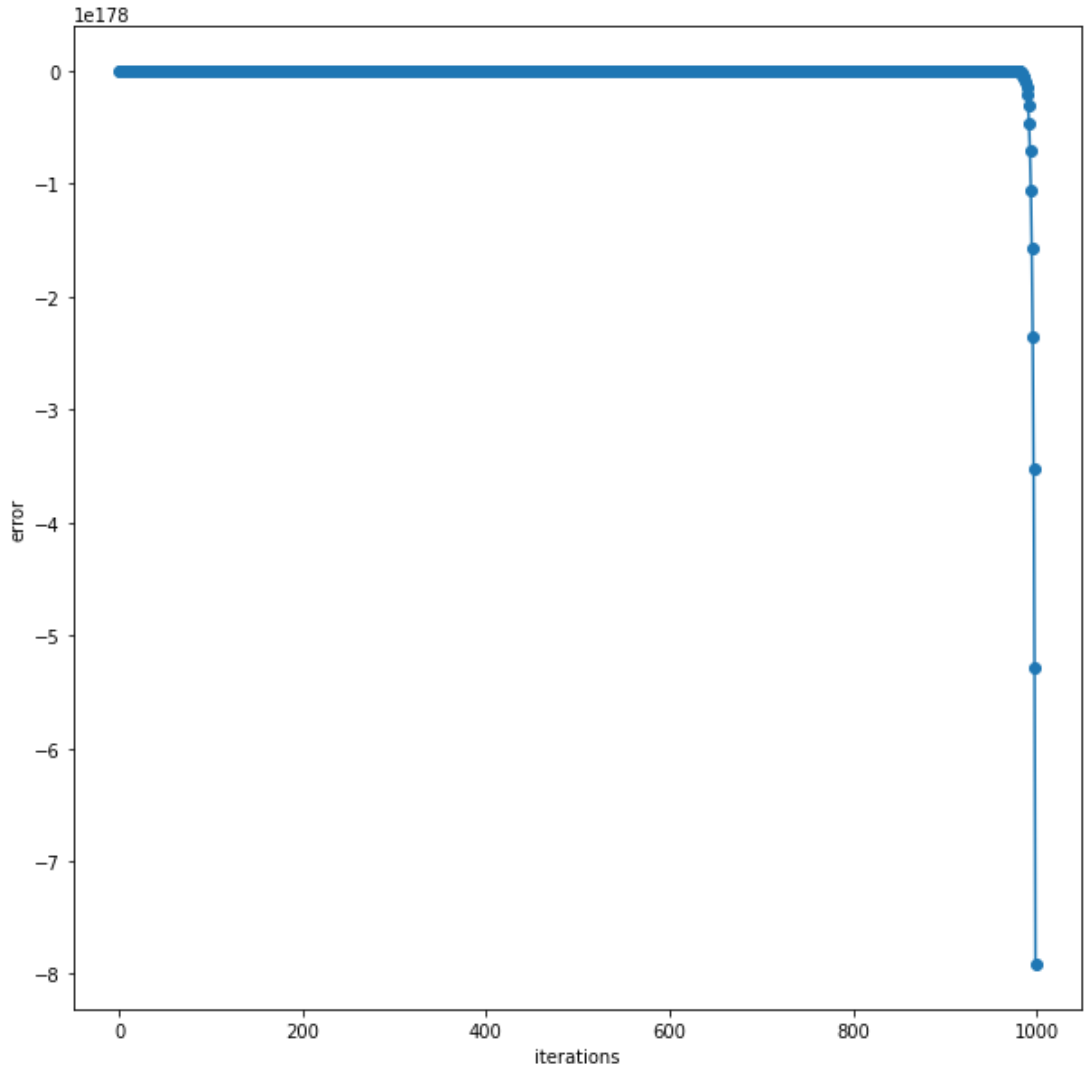


2.e

Y values vs no. of iterations



2.f



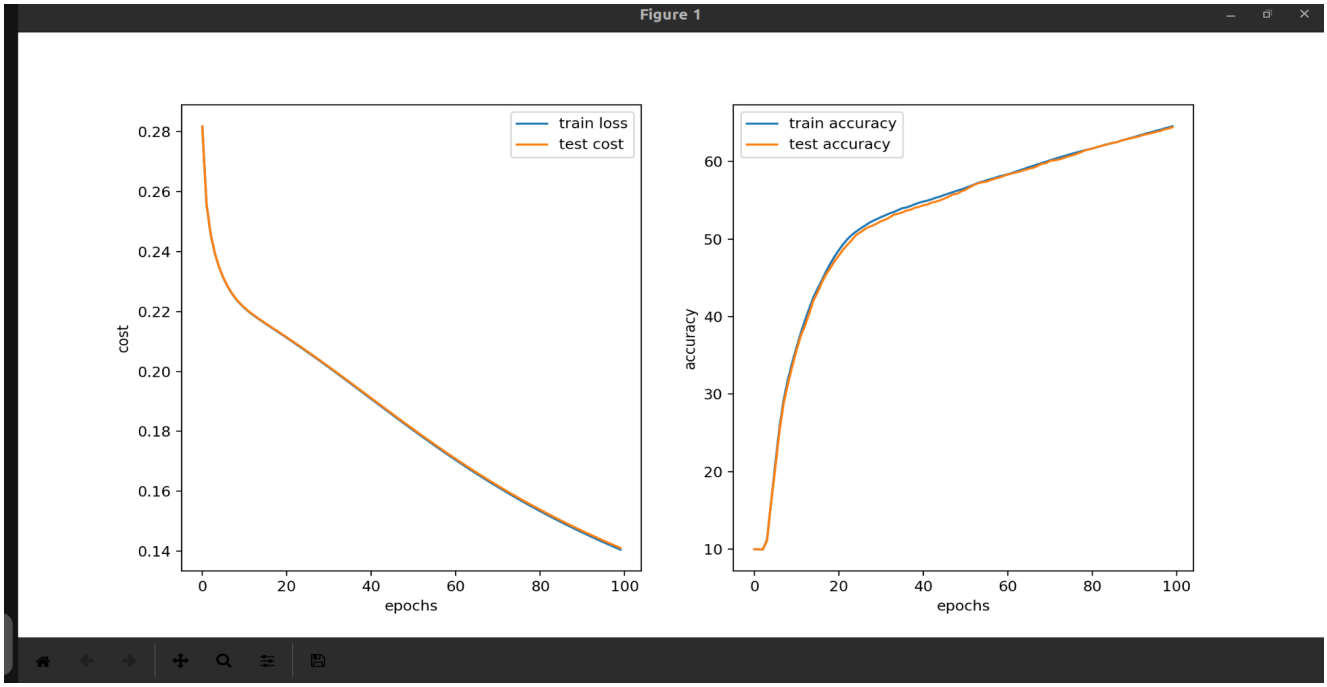
with Y=-1, we get a concave-down function which will go on decreasing upto negative infinity for higher and higher values of x1, x2 without reaching to any bottom.

Here, at later iterations, values decrease significantly hence difference between errors of initial iterations are negligible in the graph in comparison to later iterations.

Q3

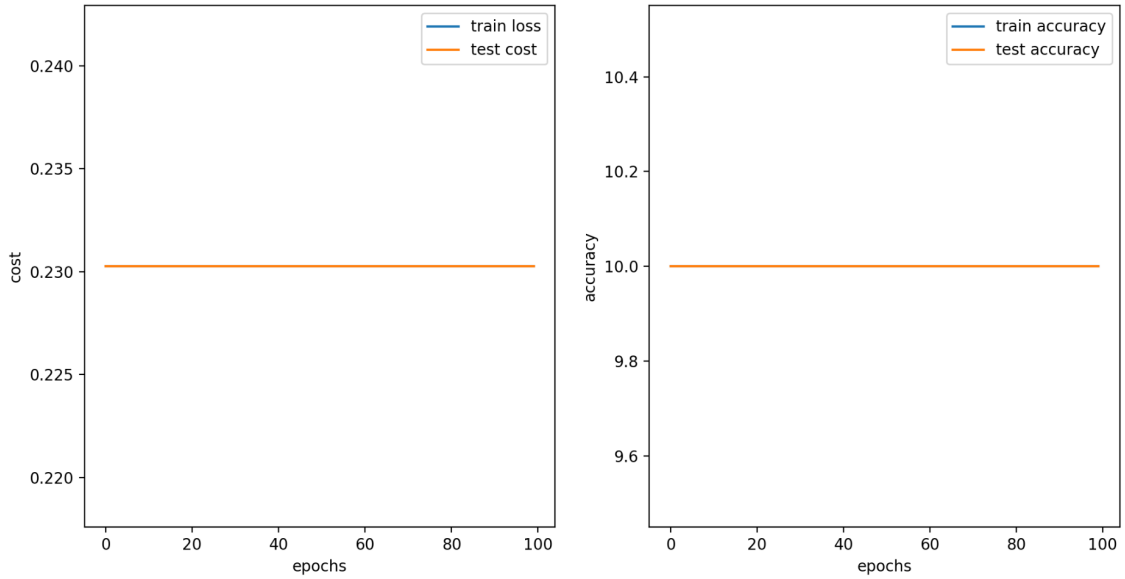
Cost and accuracy w/ epochs for train and test.

Sigmoid activation, normal distribution initialisation



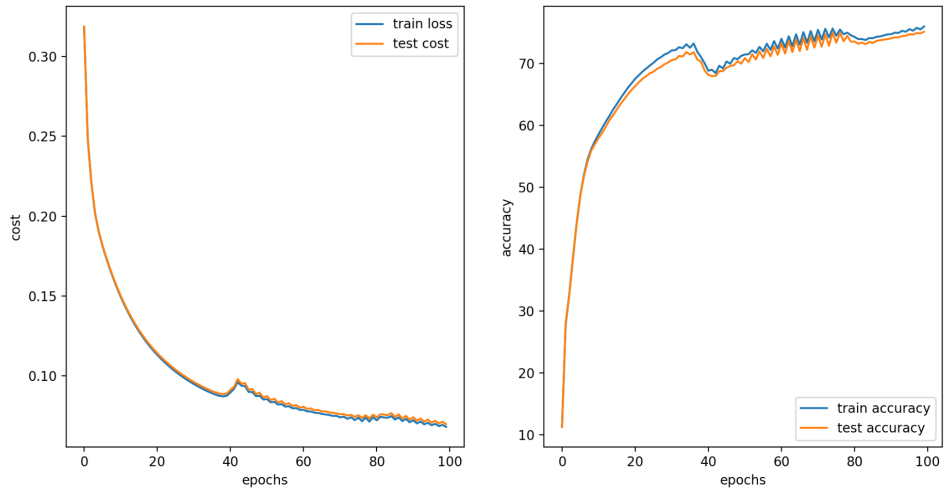
This is obvious; with progress in iterations as the weights are learned, losses decrease and accuracy improves.

Sigmoid activation, zeros initialization



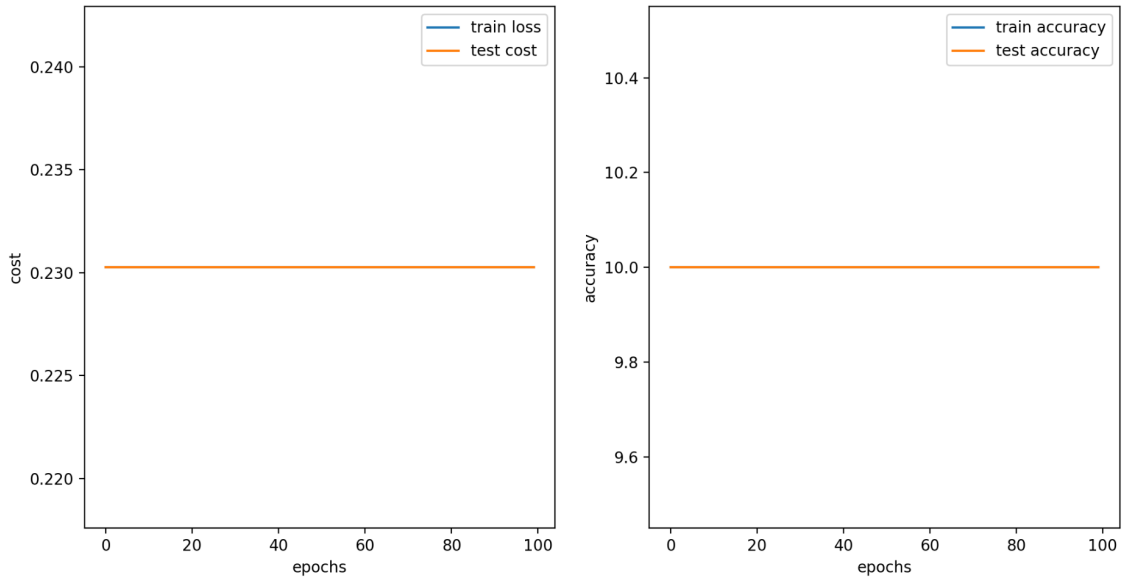
When weights are initialized to zeros, in forward pass all outputs of neurons become zero in forward pass and during backward pass, when dl/dw is calculated, this output is multiplied which makes this derivative dl/dw also zero. Hence weights would not be updated, nothing is learned.

Relu normal



We can see that both sigmoid and relu work correctly only when the initialization is non zero. This is expected as explained above.

Relu sactivation, zero initialisation



Same as for sigmoid with zero weights initialisation.