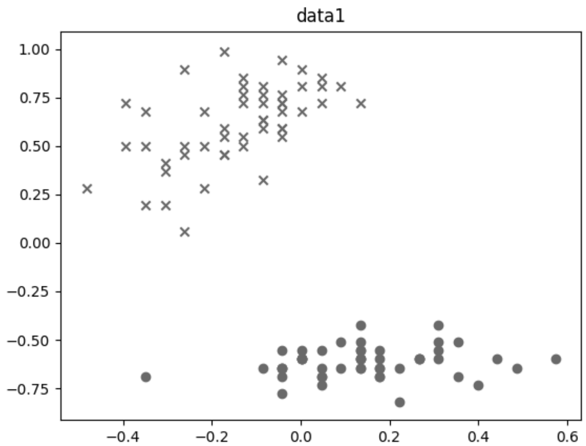
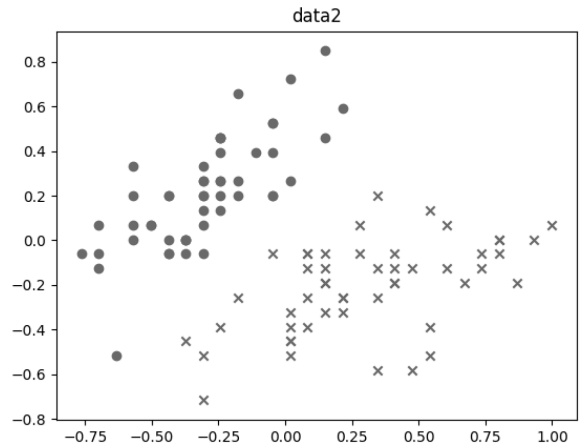
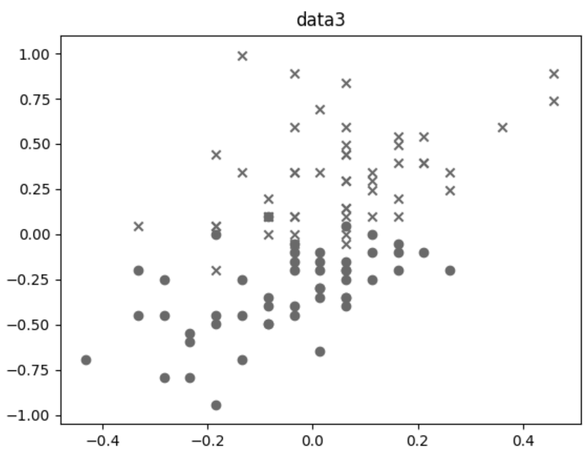
6.

(a) The plot for each dataset is shown below:





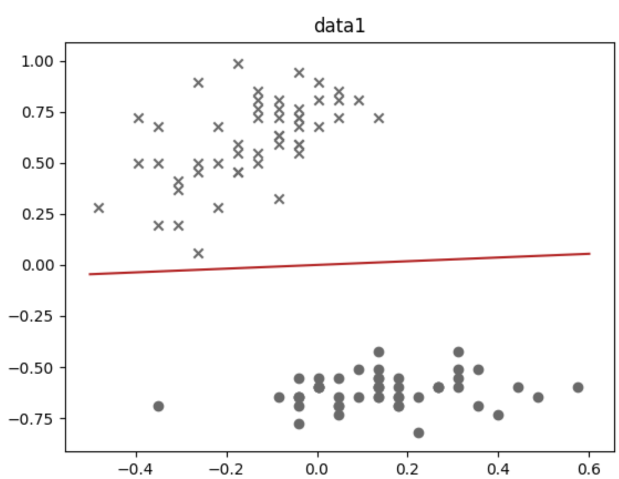


Note: o and x correspond to a datapoint whose *y* value is 1 and 0, respectively.

As you can clearly see in the figures above, data1 and data2 are linearly separable and data3 is not linearly separable.

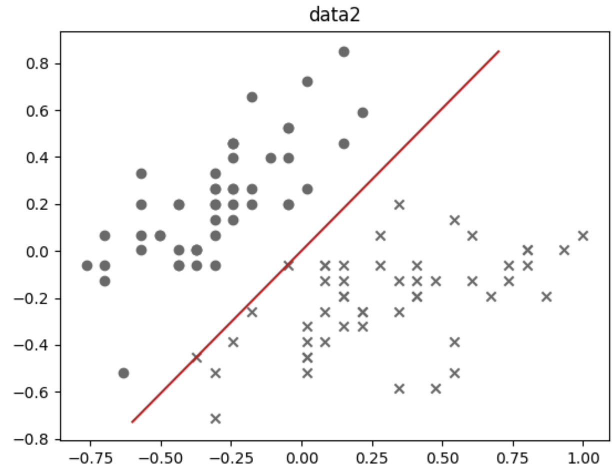
(b)

Data 1:



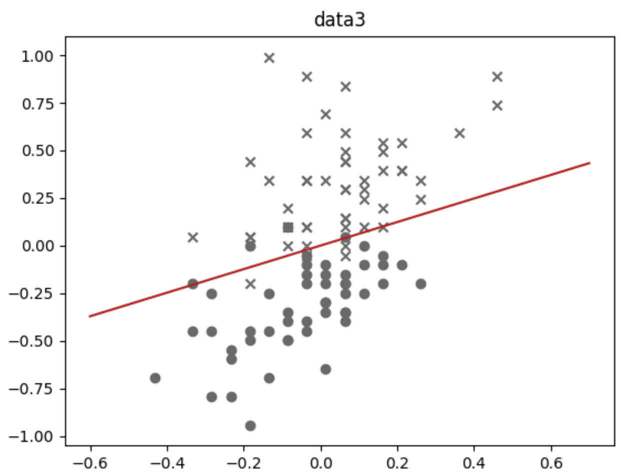

```
w =  [ 0.132014  -1.4522]
b =  0.0
u =  2
```

Data 2:




```
w =  [ -1.10918  0.91344]
b =  0.0
u =  4
```

Data 3:

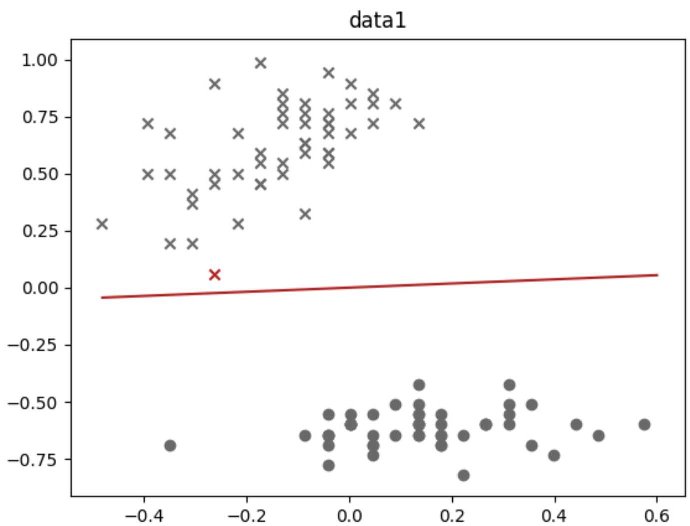



```
w =  [ 4.192936  -6.7797828]
b =  -1.0
u =  4953
```

The perceptron algorithm converges if a given dataset is linear separable. However, it does not converge (i. e. the algorithm keeps updating w till it reaches the maximum number of iterations) if the dataset is not linearly separable. Hence, for data1 and data2, the algorithm converses within MaxIter many iterations, whereas it doe not converge for data3. Moreover, the number of update for data1 is smaller than data2. This can be because data1 is intuitively easier to find a separating hyperplane than data2 because gaps between two classes ($y = 1$ and $y = -1$) are bigger than that of data2.

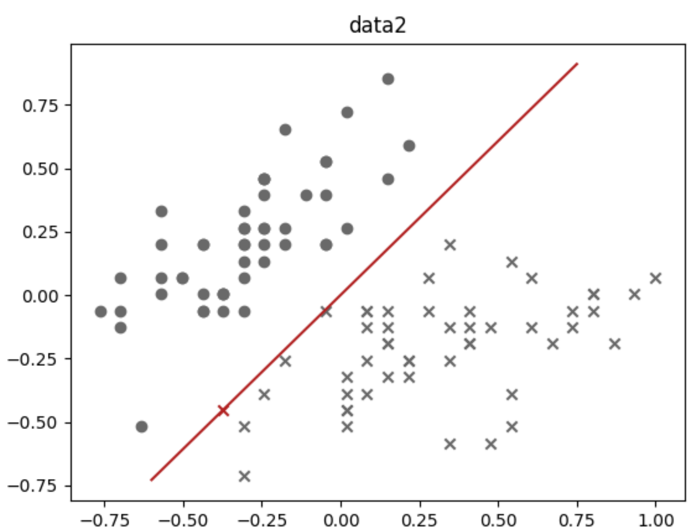
(c) Because data3 is not linearly separable, I computed $\gamma_{w,b}$ for only data1 and data2.

Data 1:




```
gamma =  0.08414391885721888
1/gamma^2 =  141.23896639191133
```

Data 2:




```
gamma =  0.0002648936222502777
1/gamma^2 =  14251382.47956429
```

For both data1 and data2, the theoretical upper bounds are much bigger than the actual numbers of update. This can be because the hyperplanes found by the perceptron algorithm are not nearly the best ones (i.e. both hyperplanes have relatively small margins compared to all other possible hyperplanes). Since $\gamma_{w,b}$ for both data1 and data2 are not at all the biggest margin for each dataset, the upper bound ($1/\gamma_{w,b}^2$) is not small enough to be a meaningful upper bound.