I)
Train accuracy = 0.833333
Test accuracy = 0.7

II)
a)
i) The learning rule is a threshold that classifies the values as 1 if prediction>0.5 or else it classifies as 0

ii) The learning rule is a threshold that classifies the values as 1 if sigmoid(prediction)>0.5 or else it classifies as 0

iii) The learning rule is a threshold that classifies the values as 1 if sigmoid(prediction)>0.5 or else it classifies as 0

b) For the Model 1 we used a learning rate of 0.1 and 30 epochs:
The first experiment gave us the results:

- Loss for Configuration 1 = 0.2833333333333333
- Train accuracy for Configuration 1 = 0.4666666666666667
- Test accuracy for Configuration 1 = 0.43333333333333335

The second experiment gave us the results:

- Loss for Configuration 1 = 0.2833333333333333
- Train accuracy for Configuration 1 = 0.4666666666666667
- Test accuracy for Configuration 1 = 0.43333333333333335

The third experiment gave us the results:

- Loss for Configuration 1 = 0.26666666666666666
- Train accuracy for Configuration 1 = 0.5
- Test accuracy for Configuration 1 = 0.4666666666666667

For the Model 2 we used a learning rate of 0.1 and 500 epochs:
The first experiment gave us the results:

- Loss for Configuration 2 = 0.21666666666666667
- Train accuracy for Configuration 2 = 1.0
- Test accuracy for Configuration 2 = 0.5666666666666667

The second experiment gave us the results:

- Loss for Configuration 2 = 0.23333333333333334
- Train accuracy for Configuration 2 = 0.9666666666666667
- Test accuracy for Configuration 2 = 0.5333333333333333

The third experiment gave us the results:

- Loss for Configuration 2 = 0.21666666666666667
- Train accuracy for Configuration 2 = 1.0
- Test accuracy for Configuration 2 = 0.5666666666666667

For the Model 3 we used a learning rate of 0.1 and 1000 epochs:
The first experiment gave us the results:

- Loss for Configuration 3 = 0.7030235590083229
- Train accuracy for Configuration 3 = 0.9666666666666667
- Test accuracy for Configuration 3 = 0.5333333333333333

The second experiment gave us the results:

- Loss for Configuration 3 = 0.708645175233855
- Train accuracy for Configuration 3 = 0.9333333333333333
- Test accuracy for Configuration 3 = 0.5666666666666667

The third experiment gave us the results:

- Loss for Configuration 3 = 0.7155311947707248
- Train accuracy for Configuration 3 = 0.9333333333333333
- Test accuracy for Configuration 3 = 0.5

For each one of the 3 Models, both the Loss, the Train accuracy and the Test accuracy were very similar, and so we can conclude that a different random initialization of weights, with a fixed learning rate, has no consequences in the results.

c)

d) Train accuracy for Configuration 1 = 0.5333333333333333
Test accuracy for Configuration 1 = 0.36666666666666664

Train accuracy for Configuration 2 = 1.0
Test accuracy for Configuration 2 = 0.5666666666666667

Train accuracy for Configuration 3 = 0.9
Test accuracy for Configuration 3 = 0.4666666666666667

We verify that both the best Train accuracy and the best Test accuracy are from the Model 2. Because the Model 2 has the best Test accuracy, we can conclude that the Model 2 is the most accurate.

The Models 2 and 3 are equivalent, since they have similar accuracies for both Training and Testing. Their sensitivities to both the weights and the bias initialization and to the learning rate are also similar.

e) By comparing the Model 3 accuracies with the Scikit-learn accuracies, we can say that the Logistic Regression Model is overting because his Train accuracy is significantly higher than the Test accuracy. The results we got from the Model 3 differ too much with the results from the Scikit-learn.