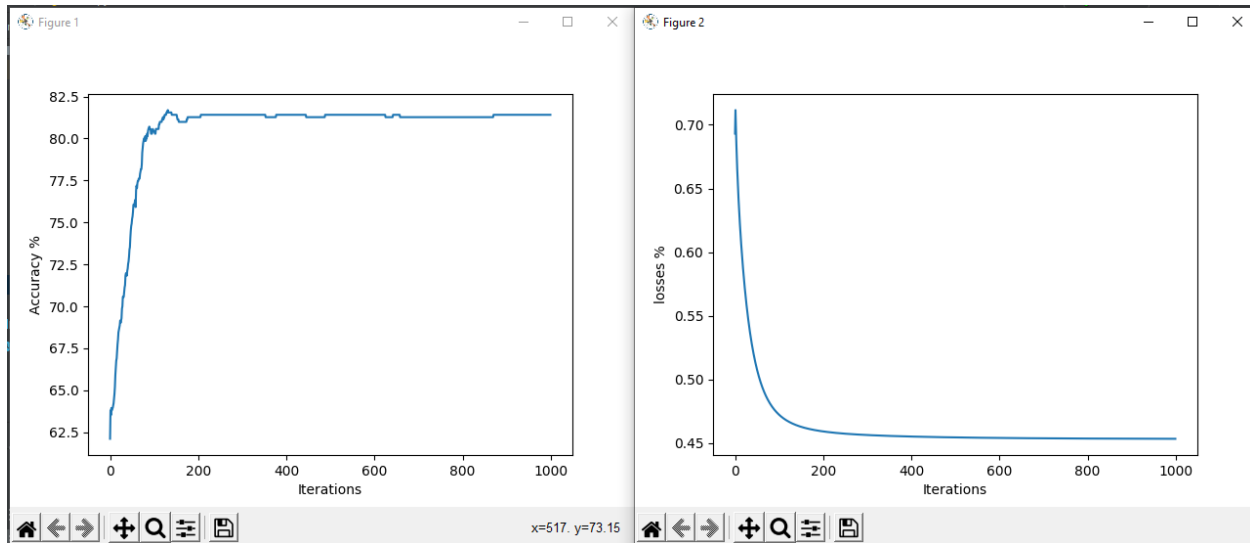Name: Ayah Hamdan

Would you survive the Titanic?

Part 1: Train a model:

1. After many tests, a good value of the learning rate was obtained to equal: 0.005.

2. For convergency, 100000 steps are required, as shown in the figure below.

Part 2: Analyze the model:

1. If the following features were used for the X variable, and by using the function logreg_inference:

Class: 1   Sex: F (1)   Age: 22   # of Siblings abroad: 3   # of parent abroad: 2
Fare: 70

Then, my probability to survive is 77.36.

2. The average accuracy of the training accuracy of the trained model is: 80.58. And it is found by using the following code:

```
print("Average accuracy: ", sum(acc)/acc.__len__())
```

3. By printing the weights (the W variable), 6 numbers are shown which represent the weight of each feature. And the values are either positive or negative numbers, where if the weight is positive, then large values of the feature, are going to push to 1 (surviving) and small values pushes to 0, while on the contrary, when the weight is negative, large values will push to 0 and small values pushes to 1. And for each feature this means:

Class (-): the higher the number representing the class, the lower the possibility to survive. (People in class number 1 are more likely to survive than in class number 3).

Sex (+): those having the 2nd feature higher (which are the females) are most likely to survive.

Age: (-): young people have more probability to survive than older people.

Siblings/spouse (-): having Siblings/spouse onboard reduces the chances to stay alive.

Parents/ Children (-): having Parents/ Children onboard reduces the chances to stay alive.

Fair (-): the more paid for the ticket, the lower the chances to live (which is not really reasonable).

* The bigger or smaller the value of the weight depends on the range of number of a feature, where wider range results in smaller weight value.

4. Survives were mostly: young, wealthy (paid more the fair and have lower classes) women, while old, poor men are most likely to die.

5. The two most influential features are the class and the sex. And to scatter plot them, small random noise was added because the features are categorical, and without the noise, the points was overlapped with each other.
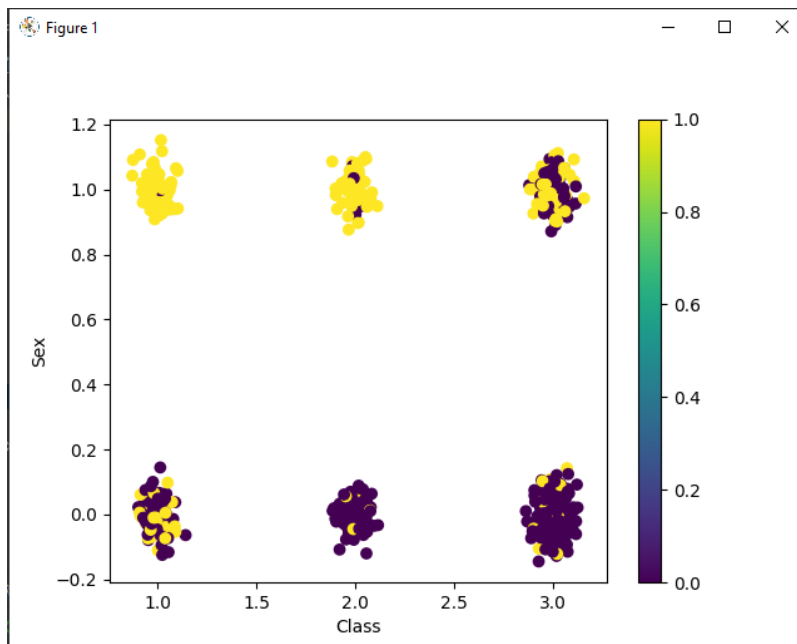
From the above figure, it can be noticed that the points representing the females are on top (level 1) while the males are in 0 level on the y-axis. And the three classes are shown in the x-axis. Also, the yellow dots represent the survived people while the purple are dead, so it can be concluded that females in the lower classes (1 and 2) are most likely to survive, while males in the 3rd are most likely to die.

Part 3: Evaluate the model:

1. The test accuracy of the model is: 77.966.

2. The model tends to be underfitting, where the results are not far from the bias.

3. To increase the performance of the model, we can work on the features, and remove useless features, by removing one feature at a time and observing the results.

In this example, trying to increase the training set by taking some sets from test set, wouldn't work because the test set is already small.

Note: I affirm that this report is the result of my own work and that I did not share any part of it with anyone else except the teacher.