Regularized Logistic Regression

- Applied to the dataset “Wisconsin Breast Cancer Database” -

Student: Benedikt Willecke
Student ID: IES19372

1. *Defined Features in the model:*

   Only the first two features are used as adding more features did not seem to improve accuracy. In fact they reduced the test accuracy. This is further explained in point 5.

2. *Train accuracy and values of cost and gradient:*

Train Accuracy: 93.60%
Cost: 0.256058
Grad:
0.072086
0.047514
0.035968
-0.257436
-0.192672
-0.102657
-3.027208
-1.996375
-1.306017
-0.880558
-25.946502
-16.099936
-11.017239
-8.131386
-6.794131
-212.032525
-127.334192
-89.953692
-70.723324
-61.787159
-61.026289
-1728.503098
-1016.001665
-734.295518
-601.506460
-548.397018
-547.161462
-593.992329

3. *Test accuracy:*
Accuracy: 98.994975%
4. *Choice of λ*
Lambda value: 9.714
With this choice of Lambda I have found a value that both maximizes the test accuracy while also minimizing the cost as explained in point 5.

5. *Efforts for improved accuracy:*
First I made a program that tried different values for Lambda and saved the value that gives the best test accuracy.

Then I tried to add one additional feature and tried to look for the optimal feature to use. I modified my program to use different features combined with different Lambda Values but unfortunately it actually decreased test accuracy to 77.88%.

I also tried to increase test accuracy by adding two additional features to the main two, that were already there but again it lead to a decreased test accuracy of 77.88%.

So I concluded that adding any additional features will not increase accuracy.

I also changed my if-statement which finds the best lambda to also consider the cost. By doing so I was able to reduce the cost from 0.400428(Lambda=0.06) to 0.389333(Lambda=0.76) which is a cost reduction of 2.77%.

I also tried reducing the value by which the program iterates through the while-loop from 0.01 to 0.001. This has helped my reduce the cost from 0.389333(Lambda=0.76) to 0.337601(Lambda=0.266) which is a further reduction by 13.29%.

By extending the search for Lambda to up to 10 I was able to find a value for Lambda that decreases the cost even further to 0.256058(Lambda=9.714).