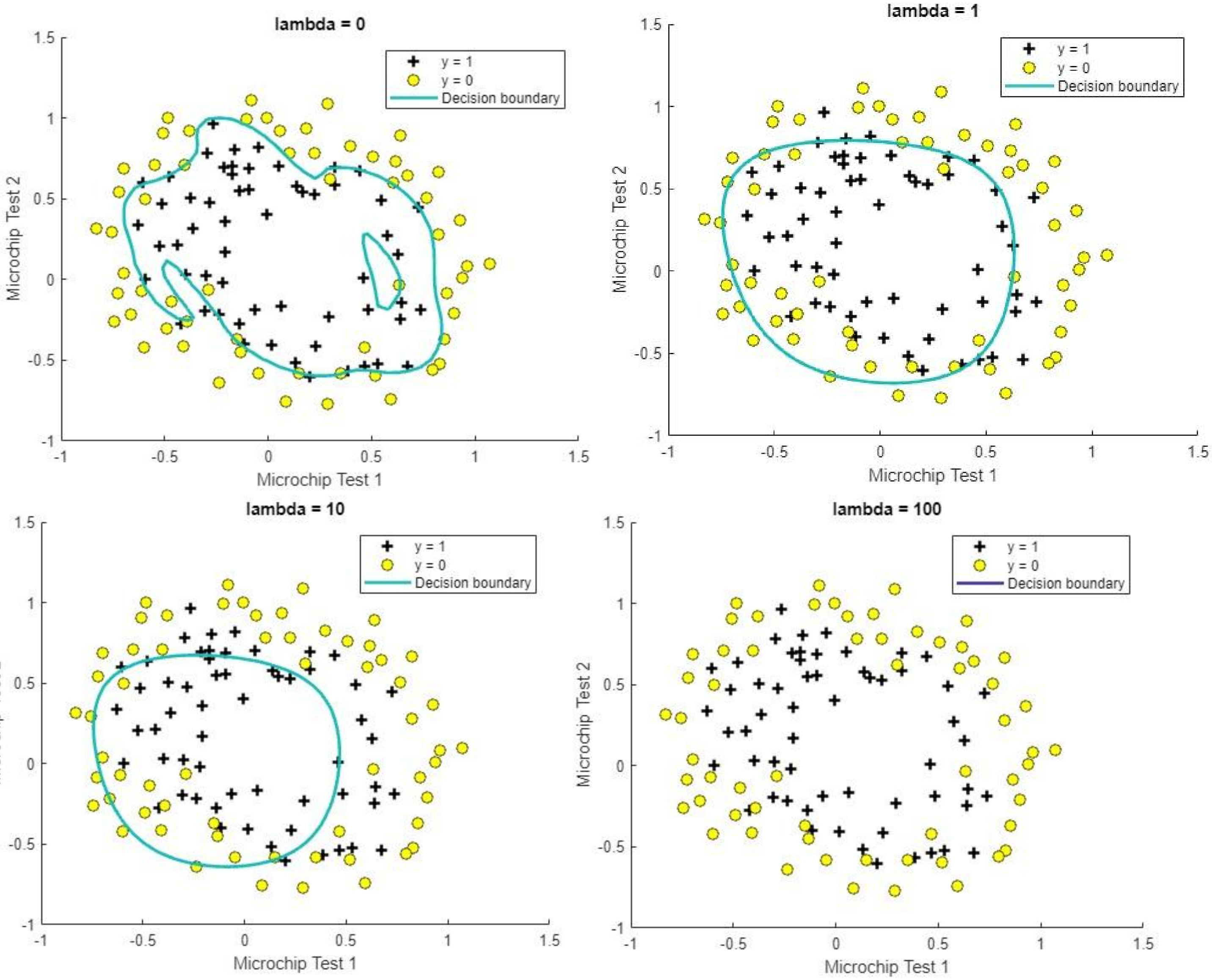
AI Homework 3 - Jaden Patel

Problem 1:



In the figures above, we can see that as the regularization parameter 𝝀 increases from 0 to 1 to 10, the decision boundary gets tighter, providing us with a simpler decision boundary that is less likely to overfit the training data. As lambda increases, the training set accuracy decreases because the model is more constrained and less able to fit the training data exactly (𝝀 = 0: 88.98, 𝝀 = 1: 76.27, 𝝀=10: 67.80, 𝝀=100: 50.85). We also observed when the regularization parameter 𝝀 is too high. In this case, 𝝀 = 100 results in no decision boundary appearing, this is likely because 𝝀 is so high that the feature coefficients have shrunk so much that they are effectively equal to 0.

Problem 3:

- C specifies the inverse of regularization strength, smaller values of C specify stronger regularization
- max_iter specifies the maximum number of iterations taken for the solvers to converge.
- n_jobs specifies the number of CPU cores to use when fitting the logistic regression model
- penalty specifies the type of regularization to use
- solver specifies the solver algorithm to use
- tol specifies the tolerance for the stopping criteria
- verbose specifies the level of output message during the model fitting process

iterations	Acc0 Pred0	Acc0 Pred1	Acc0 Pred2	Acc1 Pred0	Acc1 Pred1	Acc1 Pred2	Acc2 Pred0	Acc2 Pred1	Acc2 Pred2	Accuracy
5	15	0	0	7	10	0	9	4	0	0.555555
30	11	3	0	1	15	0	4	11	0	0.577777
50	15	2	0	3	17	0	1	6	3	0.733333

Tabulated Data for iterations 5, 30, 50

As shown in our data, increasing the number of iterations increases the accuracy of the model by improving the convergence of the optimization algorithm. However, as the number of iterations increase, the model may become too complex and overfit the training data, resulting in a decrease in accuracy. We used l1 regularization to mitigate this issue by shrinking the coefficients of less important features toward zero which reduces the complexity of the model and prevents overfitting.

References:

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

https://towardsdatascience.com/logistic-regression-and-decision-boundary-eab6e00c1e8

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

https://www.kaggle.com/getting-started/27261

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html

https://jakevdp.github.io/PythonDataScienceHandbook/02.02-the-basics-of-numpy-arrays.html

https://anandology.com/python-practice-book/working-with-data.html

https://numpy.org/doc/stable/reference/generated/numpy.ndarray.T.html

https://numpy.org/doc/stable/reference/generated/numpy.zeros.html