David Saffo

Model used:

In testing logistic regression and svm both had similar results. As a result I stuck with logistic regression as it was the model I implemented first/was the most comfortable working with.

Logistic Regression parameters:

After tweaking the parameters I could not manage to increase the performance over the default model parameters. In fact changing anything either had a negative effect or no impact. The parameters I changed were: C (inverse regularization ), penalty(type of regularization), and class_weight(weights of the class). The defaults for these parameters are 1, l2, and none respectively.

kNN:

The main parameter for kNN is the K value or the distance between points to be grouped. After running the model with the dataset several times I plotted the performance vs the k value to find the best value for k. This value ended up being around 7-9, so 9 is what I used for the k value in future tests.

Evaluation:

For each problem I tested for performance using both accuracy and f1 score. F1 score is more likely to be a better metric as it take into account not only correctness of predictions but also sensitivity in regards to bias in the data. After running both models with the test set logistic regression came out on top scoring a 76% f1 score, while my kNN implementation scored a 65% f1 score.