Yu Long Wang

Logistic Regression Regularization

The logistic regression algorithm uses the sigmoid function to map the numbers to a probability between 0 and 1.

Sigma(z) = 1/(1 + e^(-z)

For the logistic regression algorithm, it takes the features given and combines them using weights to give a probability that predicts if its malignant or benign. It makes a prediction on the training data and compares it to the actual result. It will then update weights to try and make the prediction better and repeat this process.

Example: f(x) = w1(size) + w2(texture) + w3(perimeter)

We also do this on the higher dimensions to compare if the additional combination helps with the prediction.

Second Order Example: f(x) = w1(size) + w2(texture) b+ w3(size × texture) + w4(size^2)

The regularization uses Ridge regression or L2 regularization to add a penalty to the function which is controlled by the lambda.

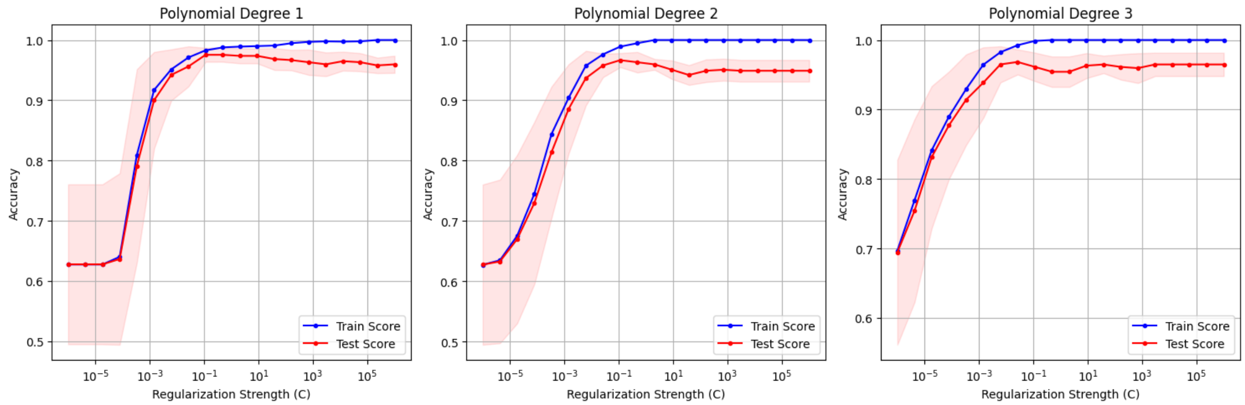
The higher the lambda the weaker the model is and the lower the lambda the stronger the regularization.

For the computation of the error measure with cross validation I split it into 5 folds, each fond with about 114 samples

Run the logistic regression for each fold of different strengths and size for training.

Eval is the sum of the folds / 5

Eval = (error_1 + error_2 + error_3 + error_4 + error_5) / 5



The overall performance for all of them for accuracy is all high around 95% and that the training score. The degree of feature transformation to the higher order did not have that much improvement in performance. Polynomial to the second degree had similar results to degree one. Polynomial degree three shows a similar result. So higher dimensions don't result in better performance. The lower the lambda the worse the performance and the higher the better usually. The best range for the lambda is from 10^-1 to 10^1 for all the degrees.

I learn that higher dimensions won't always improve performance and that the linear model with a degree of one was preferred since it simpler. I also see that the low lambdas hurt the performance more than higher lambda.