

Homework 4 Report

Name: Viprav Lipare

Student ID: 801288922

Homework Number: 4

Github Repository: <https://github.com/vipravlipare/ECGR-5105-Intro-to-Machine-Learning>

Problem 1

In Problem 1, A Support Vector Machine (SVM) classifier was used to distinguish between malignant and benign breast cancer tumors using the breast cancer dataset. The dataset was split into 80% training and 20% test sets, and all inputs were normalized. The data was split into 80% training and 20% testing sets and normalized for better performance. Four SVM kernels, linear, polynomial, RBF, and sigmoid, were tested to compare accuracy, precision, recall, and F1 score. The results were then compared to the values gotten from Homework 3, or the logistic regression method. The results were mostly similar, but slightly different.

The RBF kernel achieved the best performance with an accuracy of 98.25%, precision of 98.61%, recall of 98.61%, and F1 score of 0.9861. This means that the model correctly classified nearly all cases in the test set, hence the 98.25% accuracy. The model predicted a malignant cancer case correctly approximately 98.61% of the times it made a positive prediction, hence the precision. The model correctly identified approximately 98.61% of actual malignant cancer cases, hence the recall. All of the kernels performed well, but the RBF kernel was better than the other ones. When compared with the logistic regression model from Homework 3, which also reached 98.25% accuracy, the SVM with RBF kernel produced nearly identical results. The logistic regression version had better recall, but the RBF version had better precision, and the f1 scores for both methods were identical.

Results:

kernel: linear

Accuracy : 0.9736842105263158

Precision: 0.9859154929577465

Recall : 0.9722222222222222

F1 Score : 0.9790209790209791

kernel: poly

Accuracy : 0.9122807017543859

Precision: 0.8780487804878049

Recall : 1.0

F1 Score : 0.935064935064935

kernel: rbf

Accuracy : 0.9824561403508771

Precision: 0.9861111111111112

Recall : 0.9861111111111112

F1 Score : 0.9861111111111112

kernel: sigmoid

Accuracy : 0.9298245614035088
Precision: 0.9571428571428572
Recall : 0.9305555555555556
F1 Score : 0.9436619718309859

HW3 Results:

Accuracy : 0.9824561403508771
Precision: 0.9726027397260274
Recall : 1.0
F1 Score : 0.9861111111111112

Problem 2

In Problem 2, A Support Vector Regression (SVR) model was used to predict housing prices based on features such as area, bedrooms, bathrooms, stories, and some other features. The dataset was split into 80% training and 20% test sets, and all inputs were normalized. Three SVM kernels, linear, polynomial, and RBF. The results were plotted to compare the results with the linear regression model from Homework 2.

The RBF kernel was the best fit to the actual housing prices, as seen from the graph. The RBF line showed a smooth prediction line that followed the trend of the actual housing price data. The other two kernels were good but not as good as the RBF, the Polynomial kernel followed the trend until the larger values of the actual housing prices, and the Linear kernel did that as well but performed even worse at the larger values. The Homework 2 line was very different from the lines produced by the kernel because the line was irregular and jumped up and down on the plot, making it very hard to interpret. Overall, the RBF kernel was the best prediction of the actual housing prices, making a better line compared to the other kernels and the Homework 2 solution.

