Artificial Intelligence

Final Exam

Akshara Boppidi

Solution to 1:

1(c): Here we run the program for **1000 samples**, which implements a 10-fold cross-validation approach for finding best set of hyper parameters in a support vector machine for regression. We obtain C (penalty) = 0.125, epsilon (ϵ) = 0.0 and γ (gamma) = 2.0.

1(d): The program generates data points of the 'sinc' function contaminated with random noise. In regression we have three parameters C, ϵ, and γ. We can find best value of 'ϵ' and 'γ' with cross-validation. Smaller value of γ allows more votes around different classes. There is no great penalty and a smaller C value indicates that we want to tolerate the error to achieve margin greater. For optimal model, the C should be more and epsilon value should be less. Gamma is the parameter which handles non-linear classification.

1(c): Here we run the program for **10000 samples**, which implements a 10-fold cross-validation approach for finding best set of hyper parameters in a support vector machine for regression. We obtain C (penalty) = 0.125, epsilon (ϵ) = 0.0 and γ (gamma) = 2.0.

1(d): Here, we can see that we have the same values of 1000 samples, but there is change in test CV score. For 10000 samples, we have a better CV score than earlier. There is no great penalty, which means there is a room for errors without penalty. The C value depends on the data, when I have tried again for 10000 samples I obtained a higher value of C than before.

Solution to 2:

2(b): Here we are implementing a 10-fold cross-validation approach to find the best set of hyper-parameters C, ϵ (epsilon), and γ (gamma), in a Support Vector Machine for Regression (SVR) in the digits dataset. We obtained best value of C = 4096.0, epsilon = 1.8, gamma = 0.03125 and final validation score of 0.331107. While, we observe the plot, we can see that it has 9 colors and digits from 0-9. Here, prediction value is the colors and the digits are the true value. If we can see the digit '0' is in black color and also digit '0' is seen in purple, blue. So, we can assume that the purple '0' is really zero, but the predicted value thinks it is different digit and plots in purple. Also, we can see that we have digits 9 and 8 among the black cluster of digit zero. Here, the digit 9/8 is assumed as zero and it is plotted in black color.

2(c): We have a bigger value of C, which means we can tolerate the error and we can select more samples as support vectors. We have a great penalty and there is no room for errors without penalty. The larger c is, the less the final training error would be and takes more time for training. We have a large penalty and smaller epsilon which can cause an over fit. Other Thoughts - Finally I would say support vector machines was an interesting topic, which I enjoyed learning in class.

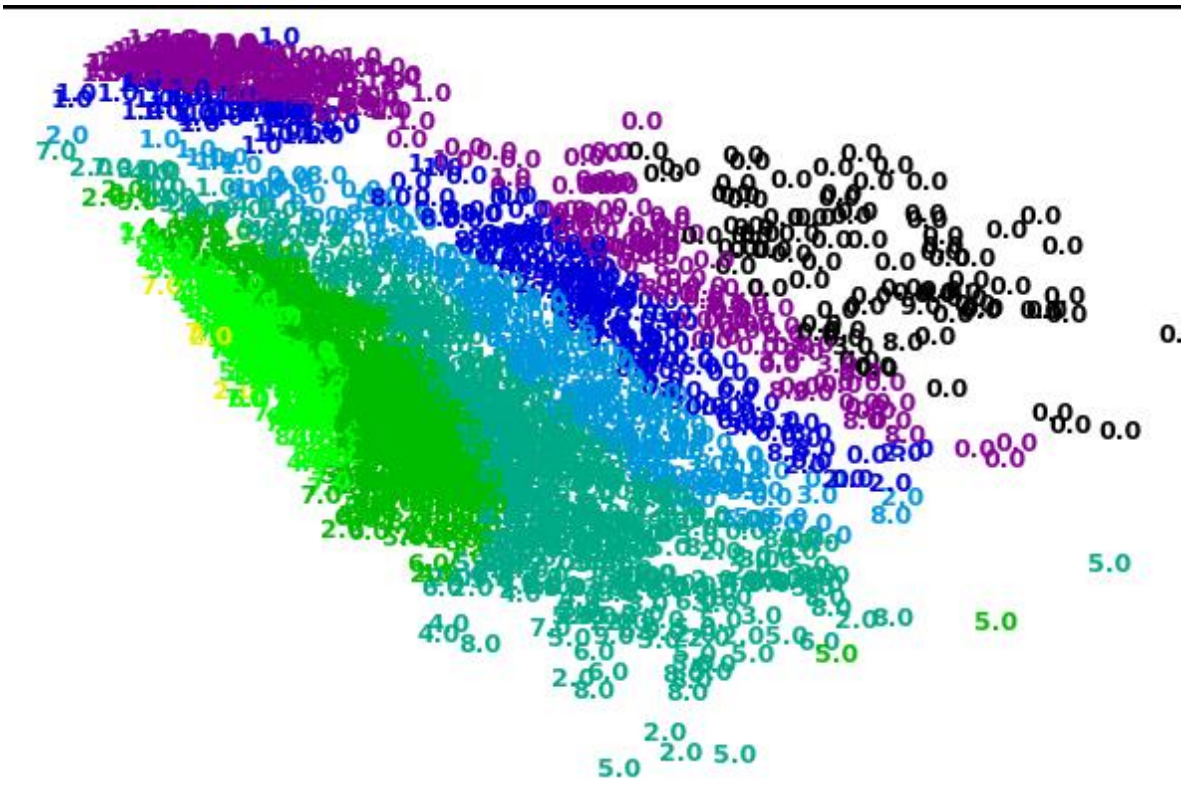

Figure: For problem 2.