Brandon Peck
Bjp9pq
Machine Learning Homework 4

2.3

Kernel	Kernel Parameter	CV Test/train Accuracy 1	CV Test/Train Accuracy 2	CV Test/Train Accuracy 3	Average 3-fold Test Accuracy
Linear	0.01	0.801/0.797	0.811/0.800	0.792/0.797	**0.801**
Sigmoid	1	0.765/0.758	0.758/0.762	0.758/0.761	0.760
Sigmoid	0.01	0.765/0.758	0.758/0.761	0.758/0.761	0.760
RBF	1	0.753/0.965	0.763/0.966	0.745/0.982	0.753
RBF	0.01	0.765/0.758	0.757/0.762	0.758/0.761	0.7604

I chose to only include three distinct kernel tests in my report. The table data is from the accuracy of the test data across three folds. The best performing model used the Linear Kernel which had an average accuracy of 0.801 with a tuning parameter of 0.01. I assume this tells us that the data can be separated relatively linearly in its k-dimensional plane. Comparing the change in accuracy in hyper parameters across the RB and Sigmoid Kernel functions, we find that the lower parameter has a higher accuracy in the RBF. This could be because a more regularized decision boundary leads to more error across the data.

2)
1)
(a)
False. Changing C will change the max-margin boundary lines. Since C is the regularization parameter, if the equation is more regularized the max-margin boundary for the support vector machine will be different. Changing the boundary can potentially change the number of support vectors because I believe a more regularized boundary can use fewer support vectors.

(b)
1, less penalty indicates wider margin in the figure. When the number of support vectors with a regularization parameter equal to zero is greater than the support vectors with a parameter equal to infinity we will see a wider margin as show in the figure.

(c)
1. (2) – This is straight forward as the second graph uses a linear decision boundary
2. (1) – This is also straight forward as the first graph has a clear polynomial decision boundary
3. (3) – The radial basis is a value that has a certain distance from the origin and manifests as circular decision boundary
1. Best (3) – The radial basis kernel is the only one with complete accuracy of dividing the data points
2. Middle (2) – The polynomial kernel has two false points above the curve which is better than three separate points in the linear kernel
1. Worst (1) – The linear kernel has three incorrectly grouped values and is the worst performing.

(d)
0. Since a support vector machine calculates based on the max-margin boundary and creates a decision boundary that has the maximum deviation from the line to a point removing a single point does not change this boundary.

2)
1)
No. If we were to draw the points on a coordinate plane we would see the positive examples diagonally placed in the first and third quadrants and the negative examples diagonally placed in the second and fourth quadrants. Since there is no line that can separate these the examples are not linearly separable.

2)
Equation given evaluates to:
$y(x)=1*w_0+x_1 w_1+x_2 w_2+x_1 x_2 w_3$
Setting w_3 to 1 shows that the max-margin will rely on the product of x_1 and x_2 which is a line across the origin which separates the positive and negative points. Since the only separating line is a line through two of the positive or negative examples, this equation satisfies that.

3)
Put a positive value at (0,2). There is no way to separate the points linearly.

4)

$K(x,x') = \phi(x)\phi(x')$

$K(x,y) = 1 + x_1x_1' + x_2x_{2'} + x_1x_{1'}x_2x_{2'}$