**Assignment2**

**Q1:**

The code is submitted in the code files section.

Chart, bar chart, histogram

Description automatically generated**Q2:**



Chart, bar chart, histogram

Description automatically generated

1. * Sample size:  
     We expected that smaller sample size will have smaller training error because there is a better chance that the sample have larger soft margin. In addition, we expected that smaller sample will have larger test error because it might not reflect the distribution as well. The results of both experiments match our expectations.
   * Training error as function of   
     Since Soft-SVM goal is to find that minimize: , we expected that larger values of will take over the hinge loss, which means that the quadradic program will prefer to minimize and ignore the error on the sample. that is why we expected that the training error as a function of will be increasing, just like we see in the second plot.
   * Test error as a function of :  
     We think that there is an optimal range for , where the test error is minimal. Below or above that range, the test error will be higher. Lower values for will cause overfitting since the hinge loss will be the major factor when calculating . Higher values for will cause underfitting since the major factor will be the soft-margin. We can see that in the plot, where the test error for both small and large samples is low when , and higher out of that range.

**Chart, scatter chart

Description automatically generatedQ4:**

1. The data set appears to be almost separable with minimal loss by a polynomial separator that is not linear in . However, there is exist where the polynomial is representable as a linear combination.
2. The minimal pair that the validation algorithm selected are: .  
   The error of the final separator over the full test data is 2%  
   All values of had the same average error over the validation test (~6%).   
   The error of the final separator over the full test data is 4%.
3. The polynomial Kernel approach achieved better result overall as we expected.
4. The polynomial kernel has larger hypothesis class compared to linear soft-SVM, therefore there is a better chance to select a separator with low validation error. However, there is also a chance to select a separator with higher validation error due to overfitting.
5. Chart

   Description automatically generatedWe used because it gave a better looking result.  
     
     
   k = 3:

Chart, histogram

Description automatically generatedK = 5:

Chart, histogram

Description automatically generatedk = 8:

1. (i)   
   where:   
     
   (ii) [ 0.01332184 0.09279347 0.0136703 0.11229866 0.00754815 0.03053518

-0.31192932 0.00444079 -0.12331285 0.00725349 -0.01331525 -0.00190763

-0.0231108 0.00041571 -0.00117185 -0.15746143 -0.00301177 -0.02398783

-0.00271744 -0.00744369 -0.02410482]  
  
(iii)

Chart, scatter chart

Description automatically generated  
  
  
(iv)

**Q5:**

1. We will show that the problem is not realizable by the hypothesis class , and therefore the minimal dependency on using PAC bounds is quadradic.

Counter example:

Let:

Suppose that is the deterministic function that decide the label over the distribution .

So, ,   
but there is no such that .  
Therefore, is not realizable by , and as we learned in class, the dependency of the sample size by is quadradic.

1. The sample size complexity is:
2. The VC dimension of is . Therefore:

**Q6:**

*Input:*

* + 1. Try
    2. If succeeded:  
       2.1. support =

2.2.

2.3.

2.4. return

3. if failed:  
 3.1 return -1

**Q7:**



where the 'th row of is

**Q8:**

1. Let , .

So:

Since is a function of , it cannot distinguish between and . Therefore, cannot return the correct value of for both vectors.   
Note that for it is possible to pad the vectors above with zeros to fit the dimension .  
Therefore, for there is not function that receive as parameter and able to return the correct value of for every , and that is why the representer theorem does not hold for the minimization problem.

1. Since the representer theorem does not hold for this minimization problem, it is not guaranteed that it is possible to calculate efficiently.

**Q9:**

1. Inner product should hold the property .

We will show counter example for :

Let .  
so:

1. As in the last question, inn Inner product should hold the property .

Let .  
so: