Andrew Elysee (aelys2)

CS446

Homework 5

1. 1. The definition of the optimization problem is   
      You can easily rationalize that is a convex function because as you move towards the 0 vector from any direction approaches 0 which means the second derivative or hessian is positive semidefinite. Also, the second derivative of is 2 which is greater than 0, so must also be convex. Making the composite function of these two convex functions, we can see that would be convex. Additionally, we can see that is a convex set. Thus, the second formulation is a convex program.
   2. If you set equal to the 0 vector in (since this is the smallest possible magnitude squared of vector ), then you can choose any and all would be valid solutions to the optimization problem since would be true. For example for and :  
      If then:  
      if then:  
      As we can see both resolve as true showing that there is no unique solution for the second formulation in this situation.
   3. The explicit constraints to the first convex program is:  
       and
   4. For :  
      For :
   5. I prefer the first version a little bit more because it feels more consistent and it is able to provide a unique solution more often than the second formulation. For me the first version is slightly more intuitive for me.
2. 1. For the hard-margin case,  
      In order to minimize , must be as close to and since can be any vector in unless is negative. If it is negative, then we have to cap it to zero. Thus, in this case , because at its best case we can produce meaning the 2 would be equal but in the worst case we produce 0 and that is the best that we can do. Therefore,   
      For the soft-margin case,  
      We can use similar reasoning in the soft margin case the hard margin case, except now we are unable to produce both negative numbers as well as numbers greater than , in other words . So since so long as , then because at its best case we can produce meaning our minimum argument would be equal to but in the worst case we produce 0 or and that is the best that we can do. Therefore,
   2. In hw5.py
   3. In hw5.py
   4. Polynomial with degree 2  
      Chart

      Description automatically generated

RBF with   
Diagram

Description automatically generatedRBF with Diagram, shape

Description automatically generatedRBF with   
Chart, diagram

Description automatically generated

1. 1. Since , we get that  
      Now that we have the gradient of our empirical risk, we can plug it into the gradient update,  
      So our final Gradient Descent update rule looks is:
   2. In hw5.py
   3. I think that looking at the plot below the logistic regression does a better job of classifying these data points because it clearly divides them into 2 separate groups while the linear classifier leaves one point on the wrong side of its boundary. With that being said I think it is important to acknowledge that if this is a sample of points, then the true population may not be linearly separable meaning that the linear classifier boundary could possibly work better in that case since it is not over correcting for the one outlier.  
      Chart, scatter chart

      Description automatically generated
2. 1. Using algebraic manipulation, we get,  
      Asserting Hoeffding’s inequality,  
      Thus we have shown that,
   2. It is important to consider that the worst-case scenario for the probability of the remaining classifiers is that they are all wrong 100% of the time meaning that . In the real world of course this would probably never happen. Since this is the worst case scenario, it will result in the highest possible probability for MAJ and thus we can use it to prove that all other possible behaviors for the arbitrary classifiers will result in a probability at less than or equal to .  
      MAJSince the last classifiers will always be wrong in the worst case we can split our summation into 2 pieces and resolve one to be .  
      Using 4a, we can reach our desired result. Thus, we have shown that, MAJ.
   3. The probability of correctly classifying x for large n is good and actually gets better the more n that we have because as n approaches infinity the exponent will approach 0.