Aaron Dardik

CS567 Problem Set 1

* 1. For test point *star* when K=4 the closest 4 points are {circle, square, triangle, triangle} so *star* is of class triangle. Note that we are assuming *star’s* coordinates are (2.2, 2.2).
  2. When K=N, *diamond* will be classified according to the most common class. As the training coordinates contain 6 triangles, 4 circles and 5 squares – diamond will be classified as a triangle.
  3. Performing N-fold cross-validation with K=1 gives 2 triangles that are correctly classified. Their coordinates are (3, 2) and (3, 2.5).
  4. KNN is a non-parametric method as we don’t fix the parameter of an underlying distribution in advance.
  5. Suppose || = |||| = |||| = 1. In this case, we have If This is equivalent to the equation Now, where is the angle between *a* and *b.* So, Now, by hypothesis we have that Now, This equals By the hypothesis that the size of we have that the above equation = Now, since By adding 2 to each side of the inequality we have By definition this means that and the result holds.

2.1) is not invertible if and only if the columns of *X* are linearly dependent. When this happens there isn’t a unique solution but rather infinitely many solutions. The solution space is a vector space of dimension D+1-N. This happens when N < D+1 as there are more variables than equations and infinitely many solutions follows as a result of having more variables than equations from the theory of systems of equations. This occurs as a result of having where and this occurs when there is a linear dependency in the columns of *X* i.e. a nonzero solution to where are the column vectors, i.e. when is not invertible.

2.2) The residual sum of squares error is We will firs take the partial derivative with respect to Next, we will calculate the partial derivative with respect to = 2By setting the derivative with respect to equal to 0, we have and by setting the derivative with respect to equal to 0, we have Now, Now, if then the latter sum in our double sum equals zero and therefore This implies that i.e.

3.1) Note that as is the optimal boundary and classifies all points correctly. So, This is equal to Therefore, Therefore,

3.2) Consider

This equals Since

Therefore we see that

3.3) From the results shown above we know that Combining these we see that We can see that by repeating this process for each mistake the algorithm makes, that if it makes M mistakes, From this we see that Also, each step we see that Now as there are M mistakes and we are increasing by at least at each mistake. Now, Combining this result with the previous result we have From this we see that and therefore the perceptron algorithm takes at most steps to converge.