Intro To ML – KNN

Amit Zulan – 207299033, Dvir Ben Zikri - 315409508

1. Code is added separately.
2. placeholder if needed.
   1. If therefore, .  
      By definition if is C-Lipschitz w.r.t. Euclidian distance then . Since both are in the sample which is sampled from , and has 0 bayes error meaning it’s deterministic, then we know and the same goes for . Therefore, we can deduce , and since . Combining everything together we get:
   2. We would like to calculate the . Meaning, for some in what is the probability . We know that . In addition, . From that we can deduce that . So, both the random prediction point and the sample point are in some ball . Since they’re both in the same ball with radius the maximum distance between is the diameter of the ball . Now we can use the C-Lipschitz property of , and we now know that both has the same label . So, the nearest neighbor prediction rule will output which we know for a fact that is the true label of due to the C-Lipschitz property. We didn’t assume anything about , so this is true . Meaning predicts the correct label. Another thing to note is that if that is closer to than that they would still have the same label due to the C-Lipschitz property. To sum up, predicts the correct label and therefore .
   3. Generally, the expected memorize error is:

The reason we can’t use the Memorize method on D and we can use on D’’ is because D’’ is a deterministic distribution and D is an indeterministic distribution. Thus, the memorization would not be able to know which indeterministic label to memorize and even if we decide on a decision rule between different labels we would always be wrong on the x we memorized that has the other label from which we memorized.

1. 1. We can see from the fig. 2 that , since these thresholds would label wrong only 2 apples from the sample .  
      .
      * , the Euclidean distance formula from .
      * will be the ERM prediction rule selected from the above . It achieves 0 empirical error on the sample S, since all blue samples are closer than 6 units from , and all red sample are further away. This is not the only function that achieves 0 empirical error on sample S. There is a range of thresholds that can achieve this.

Moreover, there is an extended hypothesis class that can also achieve 0 empirical error and might have a better prediction rule than the suggested from above.  
Choosing . This means there are more prediction rules to choose from but could reduce the approximation error. Real life reasons for this could be that the trunk is thicker or healthier in some axis.