

# Milestone 2

Andres Calderon  
acald013@ucr.edu

February 10, 2016

## 1 Introduction

This report describes some advances to accomplish the final project in the course. The main goal of the project is to perform a reliability analysis of a machine learning algorithm (kNN). This second report attempts to analyze how error in distance calculation can affect the final results and try to reason more about the error propagation using some Math. The analysis and reasoning will be supported by the previous implementation presented in milestone 1.

## 2 Error impact in the results

In  $kNN$ , when a new instance has to be classified, it looks at the distance matrix to find the closest instance to it and assign the class of that instance. If the distance between two closest instances makes an error we will miss a correct assignation. In our exercise the probability to compute a random distance is 2%, so it will be the chance to miss a correct classification. However, if we miss the instance, the algorithm will find the next closest one to assign the class. Now, the probability that two consecutive closest points have an error in their distance calculation become conditional.

Bear that in mind, one important parameter in the  $kNN$  calculation is the value of  $k$ . If we set the value of  $k$  to 3,  $kNN$  will find the three closest instances to the requested point and, according to the majority, it will assign the class. Under this consideration, error in the distance calculation has less impact in the final results due to it will have additional distance to take a decision. As mentioned before, the probability to pick  $n$  point with error in their distance calculation is low.