Model Name

Ryan Heise

June 23, 2023

1 Introduction

A K Nearest Neighbors model is one of the simplest models in machine learning. K Nearest Neighbors can be used for both classification and regression. It finds the k-closest data points in the training set to the data point whose value is to be estimated. It then uses them to infer it's value.

2 How it works

Let $x \in \mathbb{R}^n$, $X_{train} \in \mathbb{R}^{mxn}$, $y_{train} \in \mathbb{R}^m; y^{(i)}_{train} \in 0, 1,\ f : \mathbb{R}^k \to \mathbb{R}$
m: number of examples in training data, n: number of features

A K Nearest Neighbors will try and estimate the value of x. It does this by computing a distance fucntion d to each of the data points in the test set. Then it will take the top k closest points and gather their respective y values. Finally, it will then use a special function f that uses each of those y values as input to predict x.

Algorithm 1 KNN

```
procedure KNN
    for i = 1 to m do
        distances[i] = d(X^(i)_train, x)
    Set distances to be the k largest distances
    kNearestY = corresponding y values of remaining k values
    return f(kNearestY)
```

2.1 KNN for Classification

- set f to be the mode
- set the d to be the euclidean distance

For classification we find the mode of the k-nearest neighbors of the data point we are trying to predict.

2.2 KNN for Regression

- set f to be the mea of the values
- set the d to be the euclidean distance

For regression we find the mean of the k-nearest neighbors of the data point we are trying to predict.

3 Explain It Like I'm 5

KNN tries to predict a value based upon other values surrounding it. For example, consider a room with multiple colors of balls in various distinct sections of the room. If I wanted to predict the color of a ball it would make sense to look at the balls surrounding it.