

W6 Assignment – Machine Learning 1

Computational Linguistics

John Gamboa

April 22, 2022

1 k Nearest Neighbor Classifier

1.1 Behavior

Consider the image below:



Assume that the training data consists of all red and blue dots and the test data consists uniquely of the green dot. Our goal is to classify the green dot. In this context, one

could say that... (answer *TRUE* or *FALSE*)

- When $k=1$, the kNN classifier would output “red”;
answer: _____
- When $k=3$, the kNN algorithm would invariably output “blue”;
answer: _____
- When $k=3$, the kNN classifier that doesn’t take distances into account would output “blue”;
answer: _____
- When $k=6$, the kNN algorithm would output “red”;
answer: _____
- When $k=2$, the kNN algorithm would invariably output “red”.
answer: _____

1.2 Distance Measures

Calculate the Euclidean Distance between the following points:

- $a = (1, 1, 1), b = (1, 1, 10)$ answer: _____
- $a = (1, 1, 1), b = (1, 10, 1)$ answer: _____
- $a = (1, 1, 1), b = (10, 1, 1)$ answer: _____
- $a = (1, 1, 1), b = (1, 2, 2)$ answer: _____
- $a = (1, 1, 1), b = (2, 2, 2)$ answer: _____

Notice how it varies when only one dimension has changed, when two dimensions have changed, and three dimensions have changed. It may be helpful to plot these into a 3D space and see why this is the case.

Calculate the Manhattan Distance between the following points:

- $a = (1, 1, 1), b = (1, 1, 10)$ answer: _____
- $a = (1, 1, 1), b = (1, 10, 1)$ answer: _____
- $a = (1, 1, 1), b = (10, 1, 1)$ answer: _____

- $a = (1, 1, 1), b = (1, 2, 2)$ answer:
- $a = (1, 1, 1), b = (2, 2, 2)$ answer:

Notice how the variability of the Euclidean Distance is not present anymore. Can you see why this is the case?

2 Evaluation

Let's say we created a classifier that receives a new data point and produces a color indicating the class of that point. Now we want to check how well it performs for new data, that was just collected. Let's say our newly collected data had the following labels:

```
[red, red, red, blue, red, green, green, gray, red, blue, grey, red]
```

But the classifier actually produced the following labels:

```
[red, green, red, blue, red, blue, green, gray, red, grey, grey, red]
```

Calculate accuracy:

answer:

3 Bag of Words

Let's say we were given the following sentence:

```
The quick brown fox jumped over the lazy dog.
The dog was lazy and therefore didn't care about it
```

We want to create a 3-words vocabulary (plus "UNK") for this sentence, and represent it with a bag of words. Write the Bag of Words representation: (use a list of numbers, and no spaces. E.g., [3,5,2,8])

answer: