Assignment 1

Foundations of Machine Learning
IIT-Hyderabad
Aug-Dec 2021

Submitted by:
Ankita Jain
BM21MTECH14001

1. **kNN**

   (a) The training rate **increases** as 'k' increases from 1.

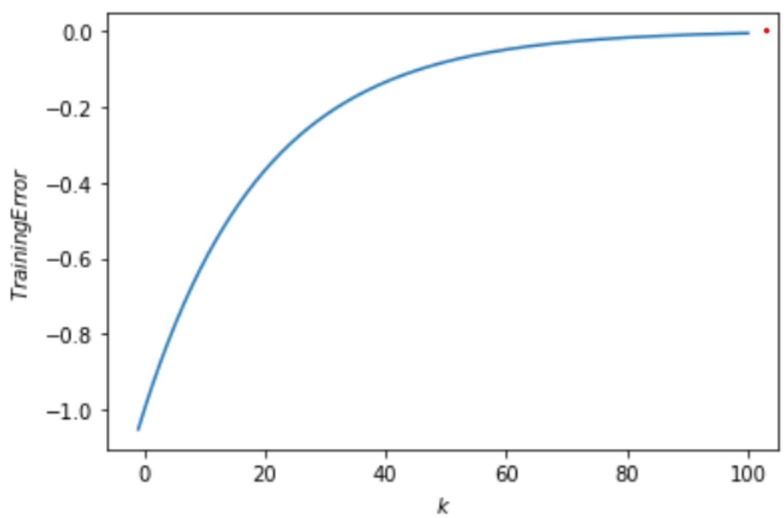

Figure 1: Training Error vs. k.

   (b) The generalisation error for k =1 is maximum because overfitting of the model has occured. As **k increases** the generalisation error reaches a **minima** after which **generalisation error increases as k increases**

   (c) The Curse of Dimensionality primarily occurs in Machine learning alforithms that are distance metric dependent some way or the other which includes k-NN

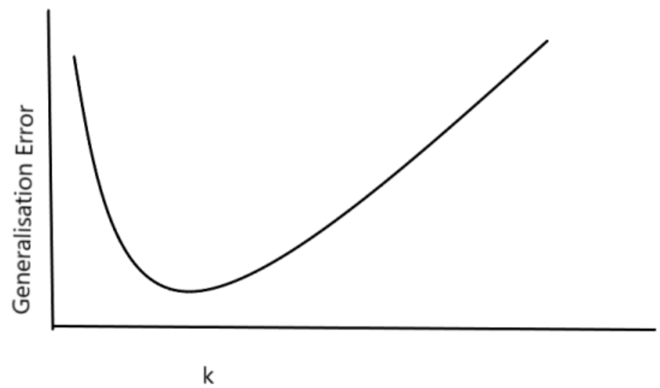

Figure 2: Generalization Error vs. k.

algorithm as well.
Two reasons supporting the above statements are as follows:

i. k-NN uses count of the closest neighbours in a given space for which it may makes use of the Euclidean distance metric among other distance metrics. Now as the number of diumensions increase the **distance metric goes on to become meaningless**.

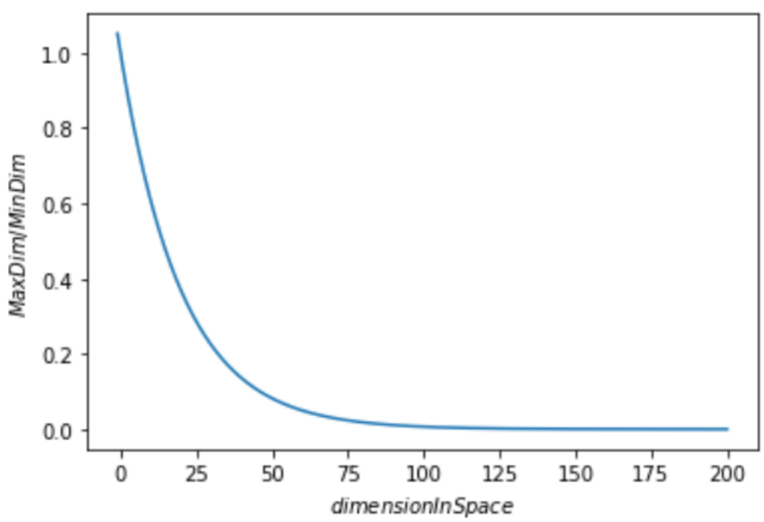

Figure 3: Max Dim/Min Dim vs. Dimensions in Space

ii. Density prer region in a given hypersurface decreases i.e we find fewer observations per region.
For example, consider a sphere inscribed within a cube assuming data points on this as dimensions increase shows that

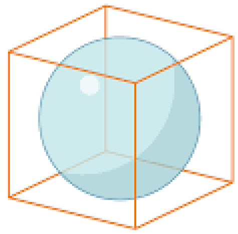

Figure 4: Cube-Sphere are regions having datapoints

$$\frac{Volumeofsphere}{Volumeofcube} \approx 0$$

Hece density reduces!

(d) The decision boundary diagram for a 1-NN is a Voronoi Diagram.

i. A decision boundary is defined to be at equal distance from the two points in question **(belonging to different classes)**

ii. Hence any point within the decision boundary is considered to be belonging to that class

Figure 5: A voromoi Diagram in

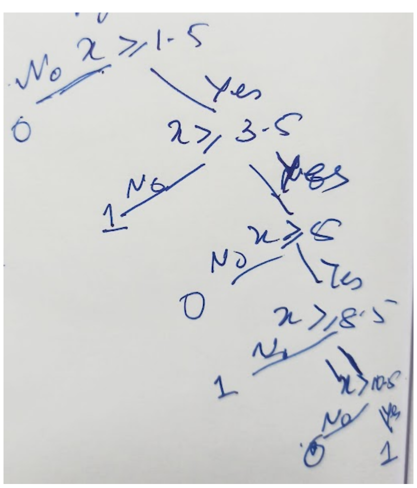

Figure 6: A Decision Tree Diagram

The decision taken on the basis of Decision Tree will alo be the same:

Hence comparing we get same results!

2. **Bayes Classifier**

Given: $\sigma_{c_1}^2 = 0.0149$, $N_1 = 10$ $\quad$ $x = 0.6$

$\sigma_{c_2}^2 = 0.0092$, $N_2 = 4$

(i) From the data in the question, using Maximum Likelihood,

$$\mu_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} x_i = 0.26$$

$$\mu_2 = \frac{1}{N_2}\sum_{i=1}^{N_2} x_i = 0.8625$$

$$P(C_1) = \frac{N_1}{\sum_{i=1}^{2} N_i} = \frac{10}{14} = 0.714$$

$$P(C_2) = 1 - P(C_1) = 0.286$$

$$P(x|C_1) = P(0.6|C_1) = 0.0675 = \frac{1}{\sqrt{2\pi\sigma_{c_1}^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_{c_1}^2}}$$

From Bayes' Theorem:

$$P(C_j|x) = \frac{P(x|C_j)P(C_j)}{P(x)}$$

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{\sum_{k=1}^{2} P(x|C_k)P(C_k)}$$

$$= \frac{0.0675 \times 0.714}{0.0675 \times 0.714 + 0.0987 \times 0.286}$$

$$\boxed{P(C_1|x) = 0.63}$$

(a)

(b) Asuming independence of attributes as in the Naive Bayes assumptio we can solve the prolem as follows":

$\Pr(Politics \mid testCase) = \frac{\Pr(Politics|testCase)\Pr(Politics)}{\Pr(test_case)}$

The Right hand side can be computed as below:

$= \frac{\Pr(1|Politics)\Pr(0|Politics)\Pr(0|Politics)\Pr(1|Politics)\Pr(1|Politics)\Pr(1|Politics)\Pr(1|Politics)\Pr(0|Politics)\Pr(Politics)}{\Pr(testCase)}$

$= \frac{\frac{2}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{5}{6}*\frac{1}{6}*\frac{4}{6}*\frac{1}{6}*\frac{6}{12}}{0.5^8}$

$= 0..38$

3. **Decision Trees**

   (a) Accuracy = 0.8119

   (b) Accuracy = 0.8281

   (c) i. Accuracy with Gini Index on an unpruned tree = 0.8119

   ii. Accuracy with Entropy on a pruned tree = 0.8265 There has been an improvement in the accuracy because of the following reasons:

   A. The only advantage of Gini index is that it is computationaly faster, there is not much effect on accuracy as we see.

   B. Pruning reduces the complexity of the tree in higher dimensions thus reducing overfitting and hence a better result on test data