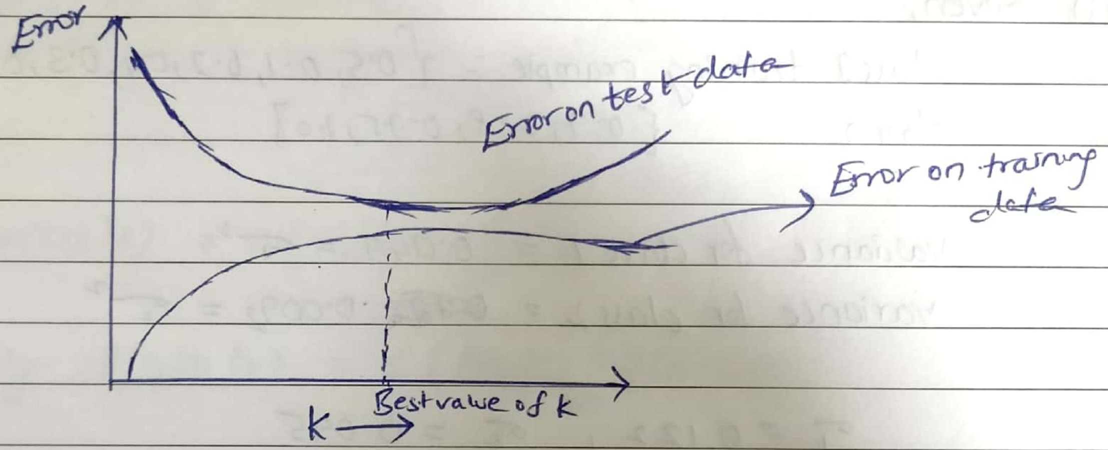
Assignment - 1

Fandations of Machine Learning

1) a) Usually, Training error gradually decreases as k value decreases (from n to 1) and as k value becomes 1, training error becomes 0

b) Graph between Error Rates and Value of k (Generalization error)



as from graph we can see that as k increases, error of test data gradually decreases where as error on training data slowly increases & finally tends to a constant value

c) In case of high dimensions, points that are drawn from a probability distribution, tend to never be close together, this breaks down the k-NN assumptions, becouse the k NN are not particularly closer than any other data points in the trainingset and also due to when dimensions in high, k NN is undesirable even due to time of computation will also be high.

d) It is not possible to build a univariate decision tree ~~(with decisions at)~~ which classifies exactly similar to a 1-NN using the Euclidean distance measure as the decision boundaries for 1-NN correspond to the cell boundaries of each point and are not necessarily parallel to coordinate axes. The decision tree boundaries would always be parallel to coordinate axes based on kinds of qns asked at each node of decision tree. To approximate a gradient by decision trees could take an arbitrary number of decisions, not possible in a decision tree.

2)

a) Given,

Class 1 training example :- {0.5, 0.1, 0.2, 0.4, 0.3, 0.2, 0.2, 0.1, 0.35, 0.25}

class 2 {0.9, 0.8, 0.75, 1.0}

Variance for class 1 = 0.049 = σ_1^2

variance for class 2 = ~~0.122~~ 0.0092 = σ_2^2

$\sigma_1 = 0.122$, $\sigma_2 = 0.095$

Distribution of class 1 = $3.27 e^{-\frac{1}{2}\left(\frac{x-0.26}{0.122}\right)^2}$

Distribution of class 2 = $4.16 e^{-\frac{1}{2}\left(\frac{x-0.862}{0.095}\right)^2}$

$P_1 = 10/\cancel{4}+10 = \frac{10}{14} = 5/7$, $P_2 = 4/10+4 = 4/14 = 2/7$

Given, test point $x = 0.6$,

$$P(\text{class } 1/x=0.6) = \frac{P(x=0.6/\text{class }1)P(\text{class }1)}{P(x=0.6/\text{class }1)P(\text{class }1) + P(x=0.6/\text{class }2)P(\text{class }2)}$$

$$= \frac{3.27 e^{-\frac{1}{2}\left(\frac{0.6-0.26}{0.122}\right)^2} \times 5/7}{3.27 e^{-\frac{1}{2}\left(\frac{0.6-0.26}{0.122}\right)^2} \times 5/7 + 4.16 e^{-\frac{1}{2}\left(\frac{0.6-0.862}{0.0959}\right)^2} \times \frac{2}{7}}$$

Scanned with CamScanner

$$= \frac{0.0482}{0.0482+0.0289}$$

$$= \frac{2}{3.2} = 0.625$$

b)

$X =$ (goal, football, gold, defence, offence, wicket, office, strategy)

given $X = (1,0,0,1,1,1,1,0)$

$$P(\text{politics}/x) = \frac{P(\text{politics}) \cdot \prod_{i=1}^{8} P(x_i | \text{politics})}{\prod_{i=1}^{8} P(x_i)}$$

$\prod_{i=1}^{8} P(x_i) \rightarrow$ as this is constant for given table, let it be 'c'

$$P(\text{politics}|x) = c\left[P(\text{politics}) \cdot \prod_{i=1}^{8} P(x_i|\text{politics})\right]$$

Similarly $P(\text{sports}|x) = c\left[P(\text{sports}) \cdot \prod_{i=1}^{8} P(x_i|\text{sports})\right]$

$\Rightarrow P(x_1=1|\text{politics}) = 2/6$, $P(x_2=0|\text{politics}) = 1/6$, $P(x_3=0|\text{politics}) = 5/6$
$P(x_4=1|\text{politics}) = 5/6$, $P(x_5=1|\text{politics}) = 5/6$, $P(x_6=1|\text{politics}) = 1/6$
$P(x_7=1|\text{politics}) = 4/6$, $P(x_8=0|\text{politics}) = 1/6$
$P(\text{politics}) = 1/2$

$$\therefore P(\text{politics}/x) = c\,\frac{1}{2}\,\frac{2}{6}\left(\frac{5}{6}\right)^4 \frac{4}{6}\left(\frac{1}{6}\right)^2$$

$$= c\,\frac{5^4 \times 4}{6^8}$$

$\Rightarrow P(x_1=1|\text{sports}) = 4/6$, $P(x_2=0|\text{sports}) = 2/6$, $P(x_3=0|\text{sports}) = 5/6$
$P(x_4=1|\text{sports}) = 4/6$, $P(x_5=1|\text{sports}) = 1/6$, $P(x_6=1|\text{sports}) = 1/6$
$P(x_7=1|\text{sports}) = 0/6$, $P(x_8=0|\text{sports}) = 5/6$

$\therefore P(\text{sports}|x) = 0$

as $P(\text{politics}/x) + \underbrace{P(\text{sports}|x)}_{\rightarrow 0} = 1$

$\Rightarrow P(\text{politics}/x) = 1$