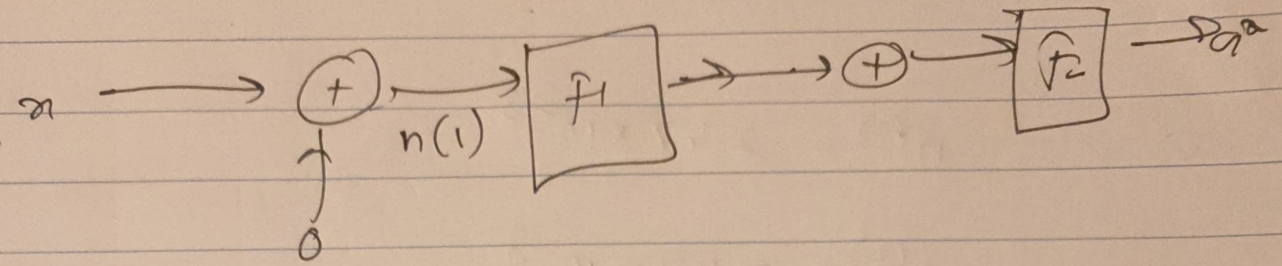
Tejas Surjit Bharambe

Q1)



$$n(1) = w^1(x) + 0$$

$$a^{(1)} = \frac{1}{1+e^{-w_1 x_1}}$$

$$\alpha = 0.2$$

$$n(2) = w^2 a^{(1)} + 0 = \frac{w^{(2)}}{1+e^{-w(1)x}}$$

$$a^{(2)} = \frac{e^{n(2)} - e^{-n(2)}}{e^{(n(2))} + e^{n(2)}}$$

$$J = (y - a^{(2)})^2$$

$x = 3$

$y = -2$

$w^{(1)} = -2$

$w^{(2)} = 1$

$n^{(1)} = -6.$

$a^{(1)} = 2.4726 \times 10^{-3}$

$n^{(2)} = 2.47 \times 10^{-3}$

$a^{(2)} = 2.47 \times 10^{-3}$

$\cancel{J = 2.47}$

$J = (-2 - 2.47 \times 10^{-3})^2$

$= (-2.00247)^2$

$J = 4.009$

$$\frac{\partial J}{\partial w^{(1)}} = \frac{\partial J}{\partial a^{(2)}} * \frac{\partial a^{(2)}}{\partial w^{(1)}} * \frac{\partial (a^{(2)})}{\partial n^{(2)}}$$

$$= 2(y - a^{(2)}) * (1 - (a^{(2)})^2) \left(w^{(2)} \left(\frac{-1}{1 + e^{-w^{(1)} x}} \right) \right)$$

$= \cancel{0.9}$

$= 0.02960$

$$\frac{\partial J}{\partial w^{(2)}} = \frac{\partial J}{\partial a^{(2)}} * \frac{\partial a^{(2)}}{\partial (n^{2})} * \frac{\partial n^{(2)}}{\partial w^{(2)}}$$

$$= 2(y - a^{(2)})(1 - (a^{2})^2) \frac{1}{1 + e^{w^{(1)} x}}$$

$= -9.902 \times 10^{-3}$

$$\boxed{w^{(1)} = w^{(1)} - \alpha \frac{\partial J}{\partial w^{(1)}} = -2 - 0.2(0.0296) = -2.00592}$$

$$w^{(2)} = w^{(2)} - \alpha \frac{\partial J}{\partial w^{(2)}} =$$

$$= 1 - 0.2\left(-9.9026 \times 10^{-3}\right)$$

$$\boxed{w^{(2)} = 1.00198}$$

$$\begin{array}{cc} & \begin{array}{cc} C & H \end{array} \\ \text{Q2)} \quad \Pi = & [0.2 \quad 0.8] \end{array}$$

(a)

$$A_1 = \begin{array}{c|c} & \begin{array}{cc} C & H \end{array} \\ \begin{array}{c} C \\ H \end{array} & \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix} \end{array} \qquad B_1 = \begin{array}{cc} & \begin{array}{ccccc} VS & S & M & L & VL \end{array} \\ \begin{array}{c} C \\ H \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.2 & 0.2 & 0.1 \\ 0.3 & 0.2 & 0.1 & 0.1 & 0.3 \end{bmatrix} \end{array}$$

$P(CH \mid VS, VL) = 0.2 \times 0.4 \times 0.1 \times 0.3 = 0.0024$

$P(HC \mid VS, VL) = 0.8 \times 0.4 \times 0.3 \times 0.1 = 0.0096$

$P(HH \mid VS, VL) = 0.8 \times 0.6 \times 0.3 \times 0.3 = 0.0432$

$P(CC \mid VS, VL) = 0.2 \times 0.6 \times 0.1 \times 0.1 = 0.0012$

Total probability of VS, VL = 0.0564

(b)

$$A_2 = \begin{array}{cc} & \begin{array}{cc} C & H \end{array} \\ \begin{array}{c} C \\ H \end{array} & \begin{bmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{bmatrix} \end{array} \quad \begin{array}{cc} & \begin{array}{ccccc} VS & S & M & L & VL \end{array} \\ \begin{array}{c} C \\ H \end{array} & \begin{bmatrix} 0.1 & 0.2 & 0.2 & 0.1 & 0.4 \\ 0.6 & 0.1 & 0.1 & 0.1 & 0.1 \end{bmatrix} \end{array} \leftarrow B$$

$P(CH \mid VS, VL) = 0.2 \times 0.4 \times 0.1 \times 0.1 = 0.0008$

$P(HC \mid VS, VL) = 0.8 \times 0.5 \times 0.6 \times 0.4 = 0.096$

$P(HH \mid VS, VL) = 0.8 \times 0.5 \times 0.6 \times 0.1 = 0.024$

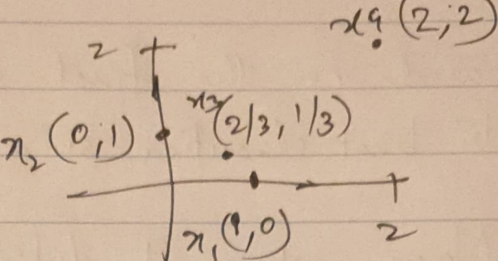
$P(CC \mid VS, VL) = 0.2 \times 0.6 \times 0.1 \times 0.4 = 0.0048$

Total probability of (VS, VL) = 0.1256

Since, the combined probability in case (b) is greater than that of case (a); ∴ the (b) case is more likely to give that sequence.

Q3) $x_1 = \begin{bmatrix}1\\0\end{bmatrix}$

$x_2 = \begin{bmatrix}0\\1\end{bmatrix}$

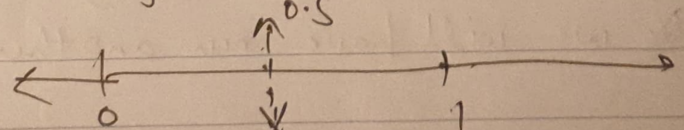


We ~~train~~ first train a classifier only based on the first feature of training vectors, then label the unlabeled vector $x_3 = \begin{bmatrix}2/3\\1/3\end{bmatrix}$.

Then we again train a classifier based on second feature of x_1, x_2, x_3.

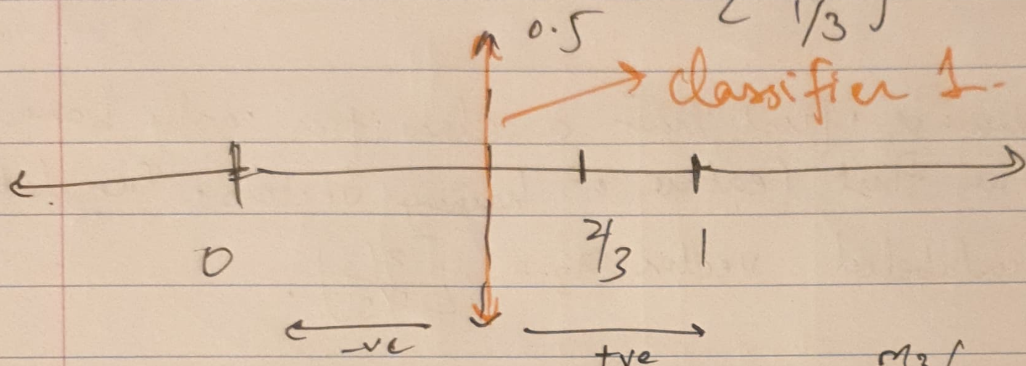
Ans) The class associated with the point $\begin{bmatrix}2\\2\end{bmatrix}$ is indeterminate if it is classified using a majority poll b/w maximum margin classifier that uses the first feature & the maximum margin classifier that uses the second feature for classification. because of a couple of reasons:

i) Training just on first feature would be equivalent to training on the below number line.



We don't know which is the best point that should be ideally used to classify it. Therefore the 1st marginal classifier will train based on the first feature and select 0.5 as the best classifier since it maximizes the distance.

Now when it labels $\begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$ based on 1st feature

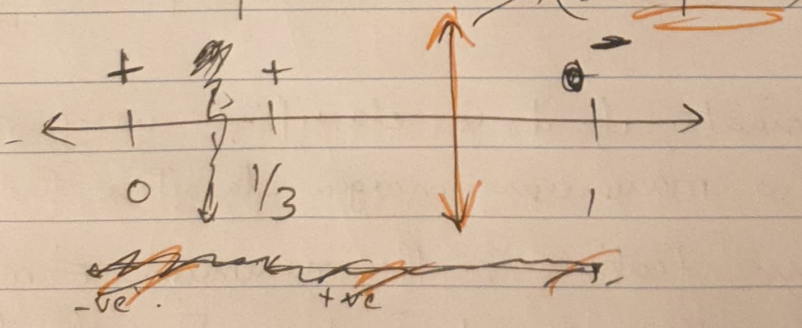


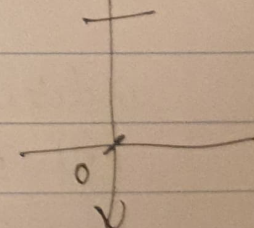
It will classify that point ($\frac{2}{3}$) to +ve class.

→ Then we train the second classifier based on the second feature: (classifier 2)

$1-\frac{1}{3}$

$\frac{2}{3}$



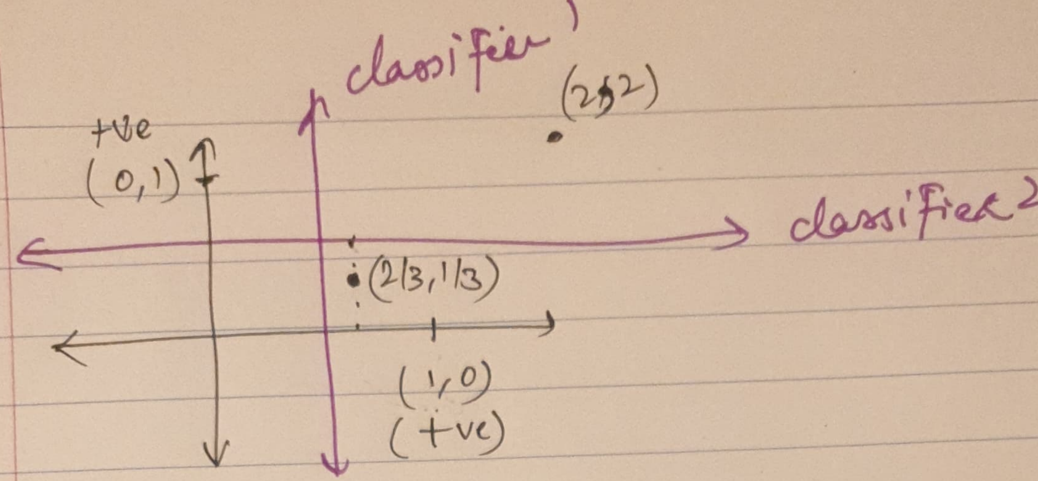


Now for the second feature and the second classifier the maximum marginal separator would lie somewhere between 1 & 1/3. i.e. 2/3

at $\frac{2}{3}$ we will have our another classifier.

Now when we classify the pt [2, 2] using majority vote,

Q3
Contd



Based on the chart above we can see that the pt (2,2) ~~is~~ will be indeterminate based ~~on~~ the majority voting from classifier 1 & classifier 2.

Classifier 1 will say its +ve class

&

classifier 2 will say its a -ve class.

There is no clear distinction or majority on which class it would be shown.

Q4) a) There is not enough information to tell.

For-eg. Suppose

$d(1,8) = 2$, $d(1,9) = 3$, $d(3,8) = 1$

$d(3,9) = 3$

$d(5,8) = 4$

$d(5,9) = 1$

The single linkage dissimilarity between $\{1,3,5\}$ & $\{8,9\}$ would be equal to 1 where ~~the~~ as the complete linkage dissimilarity would be 4.

Now, if all the inter-observation distances are equal to 2, we would have that the both single & complete linkage dissimilarities b/w $\{1,3,5\}$ & $\{8,9\}$ are equal to 2.

Hence, we can't say. & need more information.

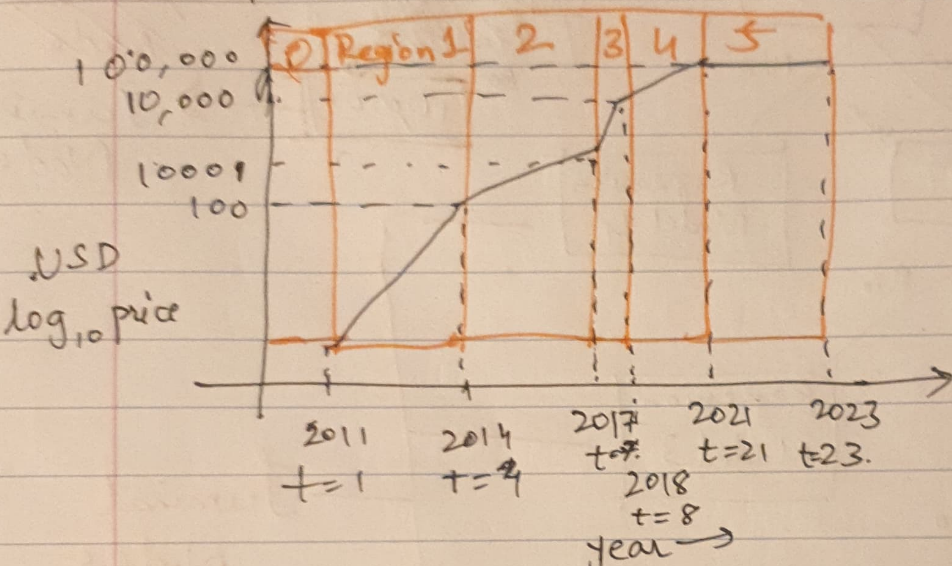
(b) They would fuse at the same height for both single & complete linkage problems.

Suppose if $\{8,9\}$ is 3, then both single & complete linkage ~~would~~ dissimilarities would be 3.

So, they both would fuse at height 3 for single & complete linkage

Q5) Dependent variable = $\log_{10}$ price.

Independent variable = time, $t \in [1,13]$

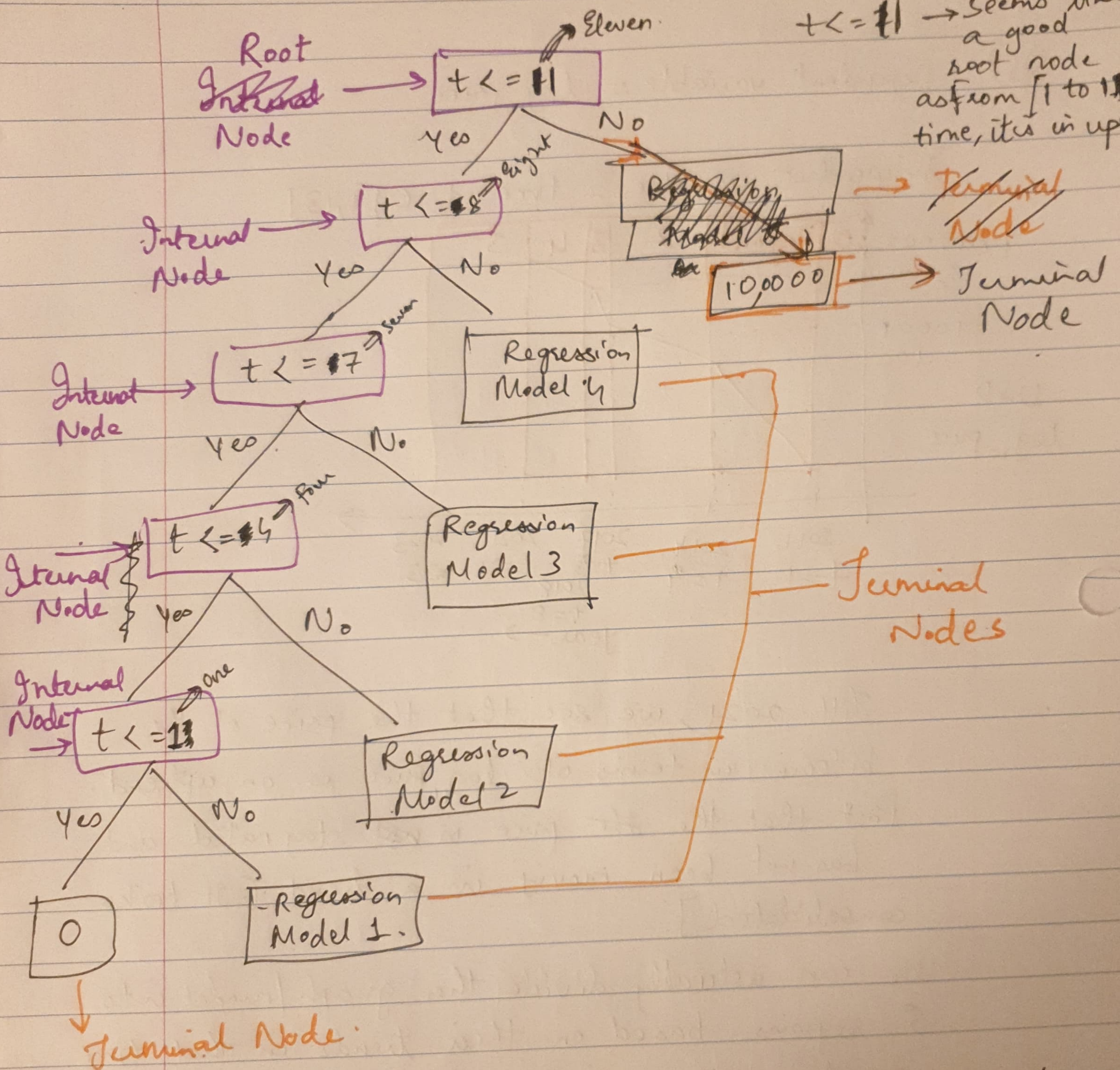


Till 2021, we see that the price of the bitcoin in terms of $\log_{10}$ price is on uptrend. Post that the ~~the~~ price is ~~not~~ stagnated and has not been moving in a trend. [It looks consolidated].

We can actually divide the graph/model into 5 regions based on their trends in that time period, as shown in the diagram above.

Now, in each region, we can fit a regression model that would predict the price / y (USD) based on the independent variable. i.e. time. except the 5^{th} Region & 0^{th} region where the output ~~be~~ would be constant of 10,000.0 & 0 respectively.

We can fit a decision tree like.



t<=11 → Seems like a good root node as from [1 to 11] time, it's in uptrend

Note that all the regression models would output the y i.e. $\log_{10}$ price using the independent variable of time 't'.

$\boxed{t<=11}$ Seems like a good root ~~note~~ node as from [1 to 11] time period, it is uptrend & post that it went into ~~dow~~ consolidation.