Machine learning

Presented by

Prof. Pushkar Sathe

Machine learning (ML)

- Subset of Artificial Intelligence (AI)
- Works in a similar way to human learning
- Ability of system to independently find solution
- Ability to automatically learn and improve from experience
- No explicit instructions used
- Builds a mathematical model based on sample data (training data)

ML applications

- Virtual Personal Assistants e.g. Alexa
- Face recognition on Facebook
- Video recommendation of YouTube
- Google translator
- Spam detection on Gmail
- Medical field : diagnosis and prognosis
- Sentiment analysis

ML algorithm types

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised learning algorithms

- Simple linear regression
- Multiple linear regression
- Logistic regression
- Support Vector Machine
- Decision tree
- Random forest
- Naïve Bayes

Simple linear regression

- Statistical method used for regression
- Study of relationship between independent and dependent variables
- Involves one dependent variable and one independent variable
- Both the variables are continuous in nature
- Model is ‘best fitting line’

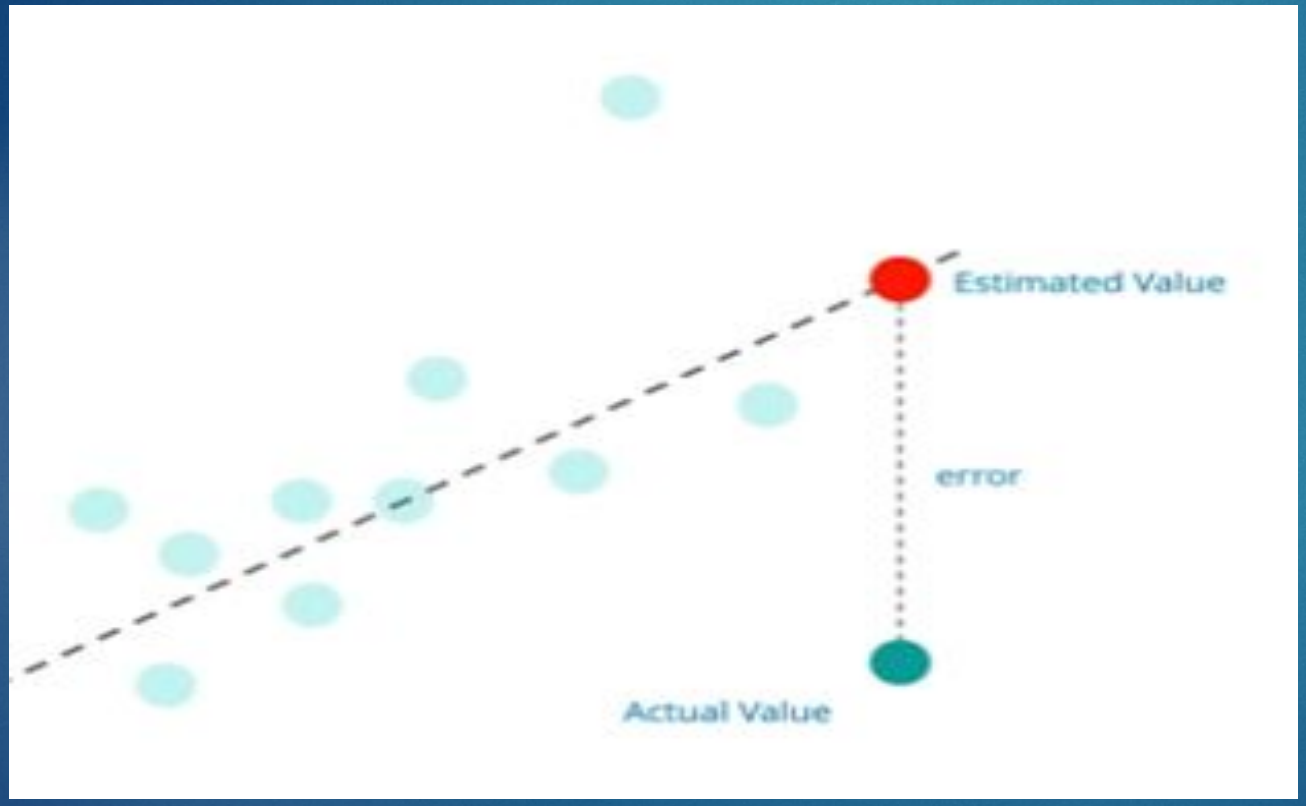
Example


Salary vs Experience (Test set)
120000
100000
80000
60000
40000
Salary
2
4
6
8
10
Years of Experience

Best fit line

- $y = m.x + c$
- Line belonging to optimum values of ‘m’ and ‘c’ is best fit line
- Least square approach is used
- Error = (predicted value – actual value) of dependent variable
- Sum of squares of errors of all data points should be minimum

Error



Multiple linear regression

- Involves several independent variables and single dependent variable
- Equation describing the relationship is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$$

where, for $i = n$ observations:

y_i = dependent variable

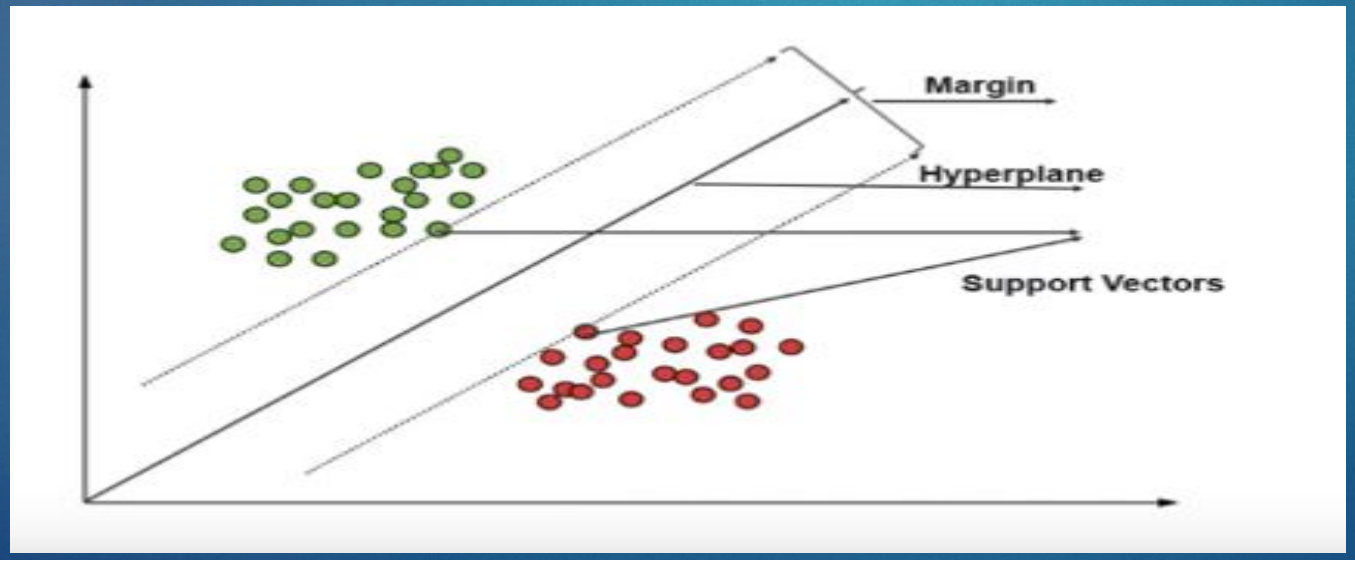
x_i = expanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

Support Vector Machine (SVM)

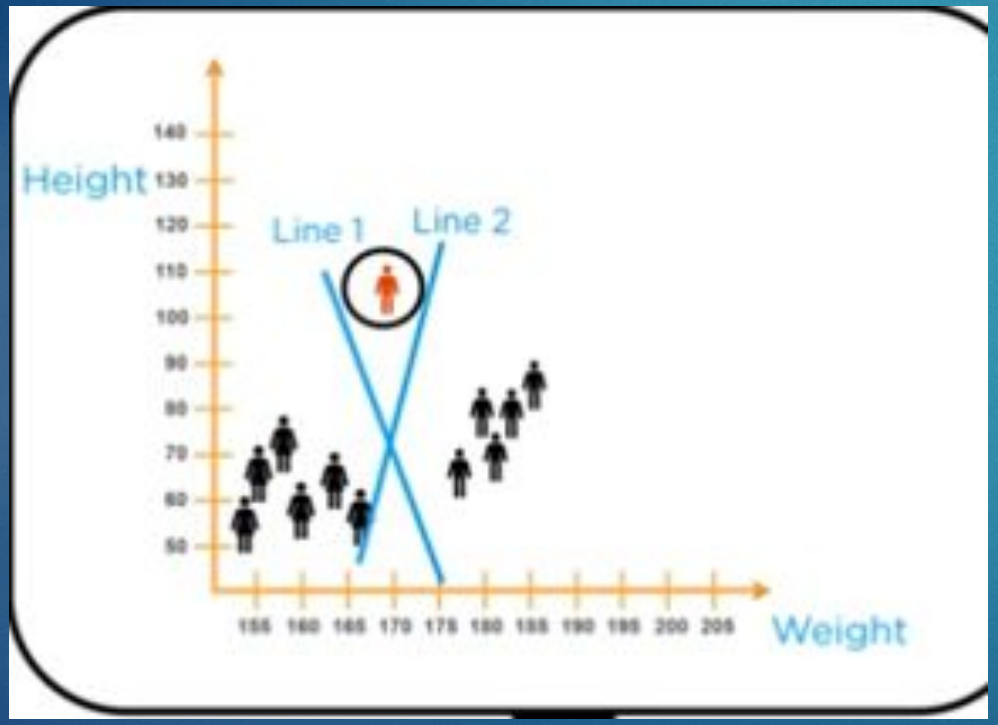
- Discriminative classifier works on separation of hyper-planes
- Very effective even when data is high dimensional



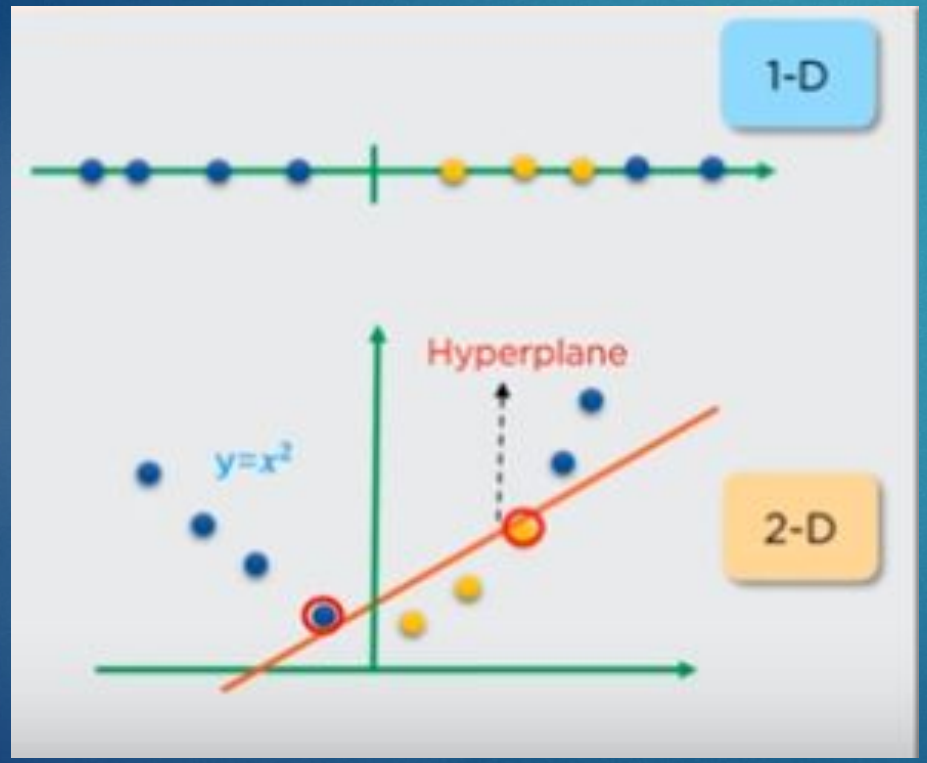
Working of SVM

- Forms hyper plane between data points
- Margin is a distance between nearest data point and hyper plane
- Hyper plane giving highest margin should be selected
- Forming hyper plane is difficult in case of non linear or inseparable problem
- Use of Kernels : Transformation to higher dimensions

Example : SVM



Example : Kernels in SVM



Types of kernels

Kernel type is selected based on distribution of data points

- Linear
- Polynomial
- Radial Basis Function (RBF)

Decision tree

- Can be used for classification as well as regression
- Algorithm involves representation of data as a tree
- Drawn upside down with its root at top
- At every node one branch is selected based on feature value
- Algorithm is used especially in decision making

Example

Age	Competition	Type	Profit
Old	Yes	S/W	Down
Old	No	S/W	Down
Old	No	H/W	Down
Mild	Yes	S/W	Down
Mild	Yes	H/W	Down
Mild	No	H/W	Up
Mild	No	S/W	Up
New	Yes	S/W	Up
New	No	H/W	Up
New	No	S/W	Up

Formulae

$$I.G. = \text{Entropy}(\text{complete set}) = \frac{-P}{P+N}\log_2\left(\frac{P}{P+N}\right) - \frac{N}{P+N}\log_2\left(\frac{N}{P+N}\right)$$

$$\text{Entropy of attribute} = \sum_{i=1} \frac{P_i + N_i}{P+N} I(P_i, N_i)$$

$$\text{Gain} = I.G. - E(A)$$

Solution

Here P = No. of downs = 5

N = No. of ups = 5

$$\therefore \text{I.G.} = -\frac{5}{10}\log_2\left(\frac{5}{10}\right) - \frac{5}{10}\log_2\left(\frac{5}{10}\right)$$

$$= 1.$$

(If there are 2 no.s & both are same like in this case $\frac{5}{10}$ & $\frac{5}{10}$, then I.G. = 1).

Solution

To find Entropy & Gain of attributes

i) Age

	Down	up
old	3	0
mid	2	2
new	0	3

$$E(A) = \frac{3+0}{5+5} I(3,0) + \frac{2+2}{5+5} I(2,2) + \frac{0+3}{5+5} I(0,3).$$

$$I(3,0) = -\frac{3}{3+0} \log\left(\frac{3}{3+0}\right) - \frac{0}{3+0} \log\left(\frac{0}{3+0}\right)$$

$$= 0$$

(Whenever out of 2 nos, one number is zero $I = 0$)

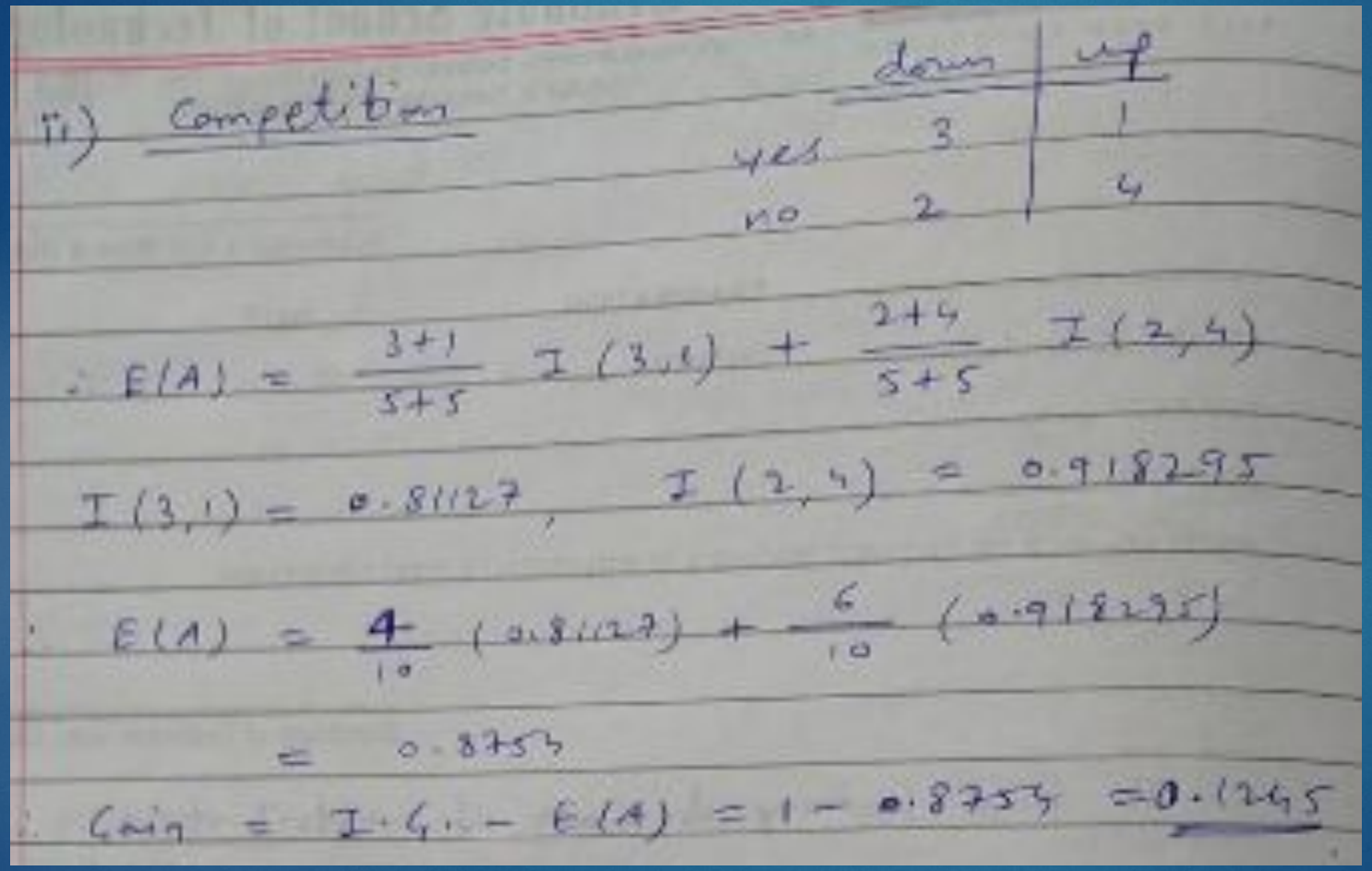
& $I(2,2) = 1$ (∵ both nos are same)

$$\therefore E(A) = \frac{3}{10}(0) + \frac{4}{10}(1) + \frac{3}{10}(0)$$

$$= 0.4$$

$$\therefore \text{Gain} = I.G. - E(A) = 1 - 0.4 = \underline{0.6}$$

Solution



ii) Competition

	down	up
yes	3	1
no	2	4

$$\therefore E(A) = \frac{3+1}{5+5} I(3,1) + \frac{2+4}{5+5} I(2,4)$$

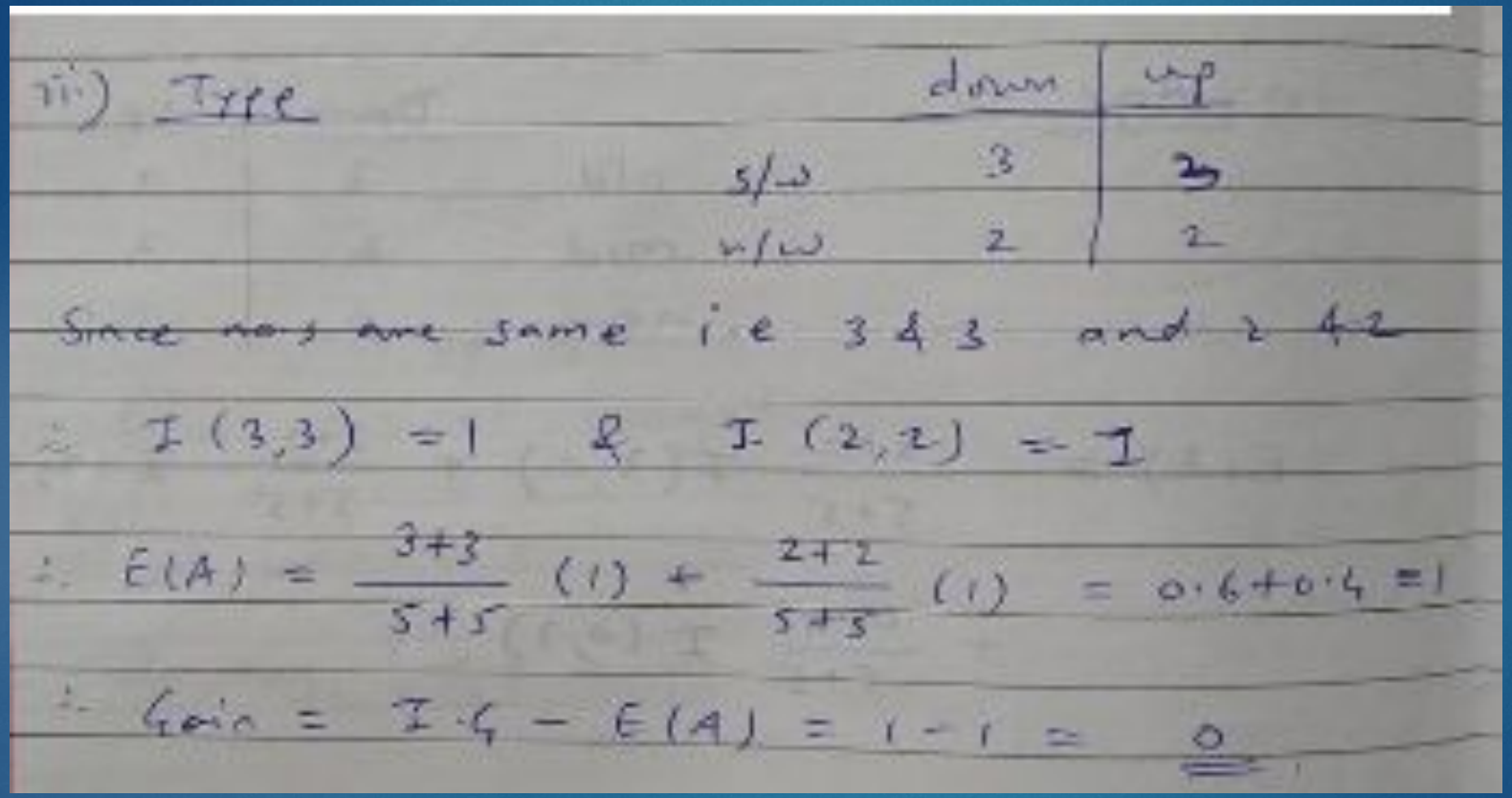
$$I(3,1) = 0.81127, \quad I(2,4) = 0.918295$$

$$\therefore E(A) = \frac{4}{10}(0.81127) + \frac{6}{10}(0.918295)$$

$$= 0.8753$$

$$\therefore \text{Gain} = I.G. - E(A) = 1 - 0.8753 = 0.1245$$

Solution



iii) Type

	down	up
s/w	3	3
n/w	2	2

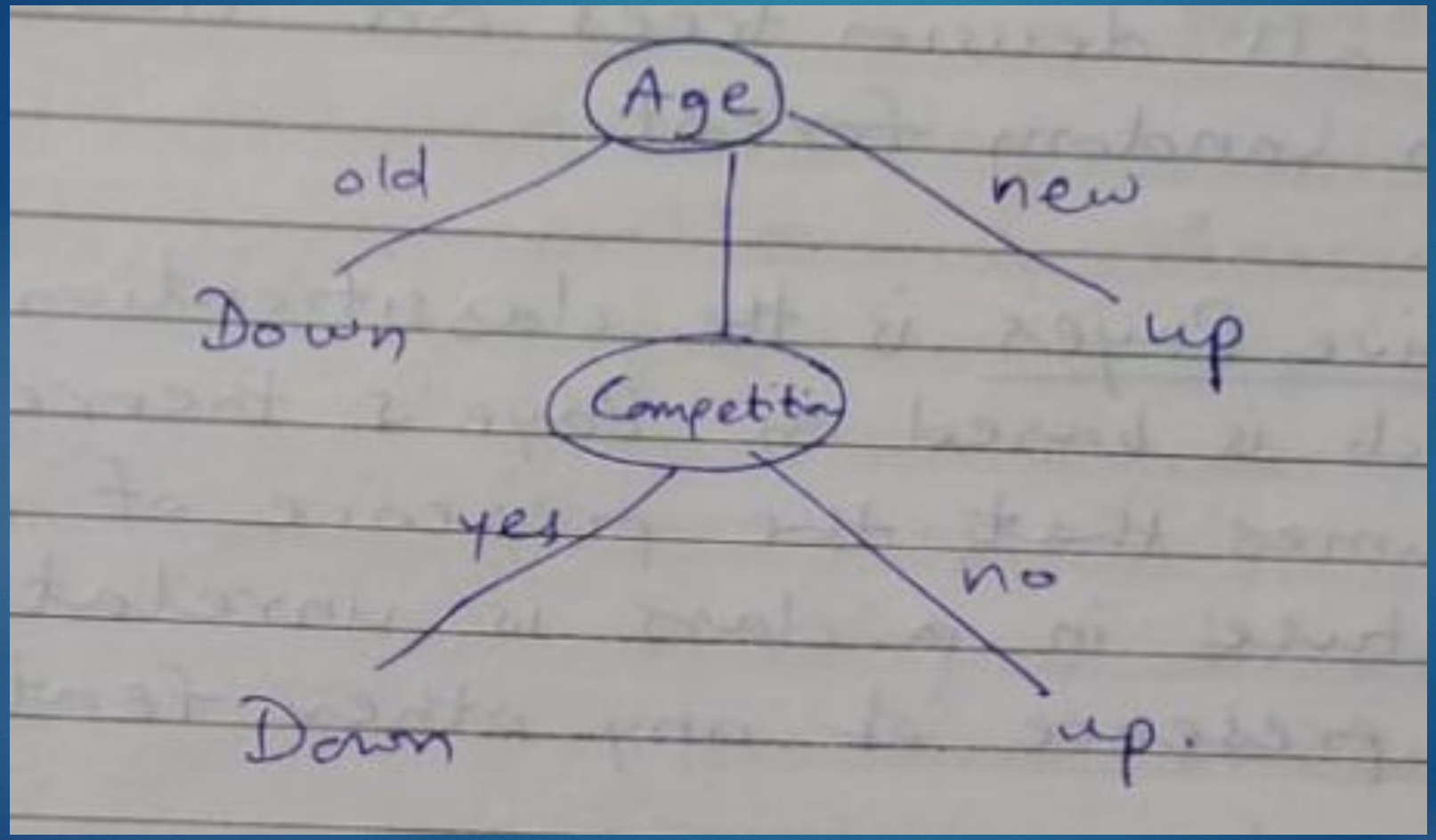
Since no's are same i e 3 & 3 and 2 & 2

$\therefore I(3,3) = 1$ & $I(2,2) = 1$

$\therefore E(A) = \frac{3+3}{5+5}(1) + \frac{2+2}{5+5}(1) = 0.6+0.4 = 1$

$\therefore Gain = I.G - E(A) = 1 - 1 = 0$

Solution



Random forest

- Concept is based on wisdom of crowds
- Consists of large number of individual decision trees
- Trees should have low correlation among each other
- Each tree gives out a class prediction
- The class with most votes becomes model's prediction

Thank you