

14/22

IAT - II

M. Sanker Abinash

Data mining

1912098

1. Adaboost, Boosting Algorithm

2. K means Partitioning Algorithm

CO1 - 7

CO2 - 5

CO3 - 1

1.) Adaboost Algorithm

Verified
M. Sanker

* This algorithm is used to the weight of the point at a time.

* Identify the correct mistake of the point

* Increase the number of missing point and identify the correct mistake of the missing point.

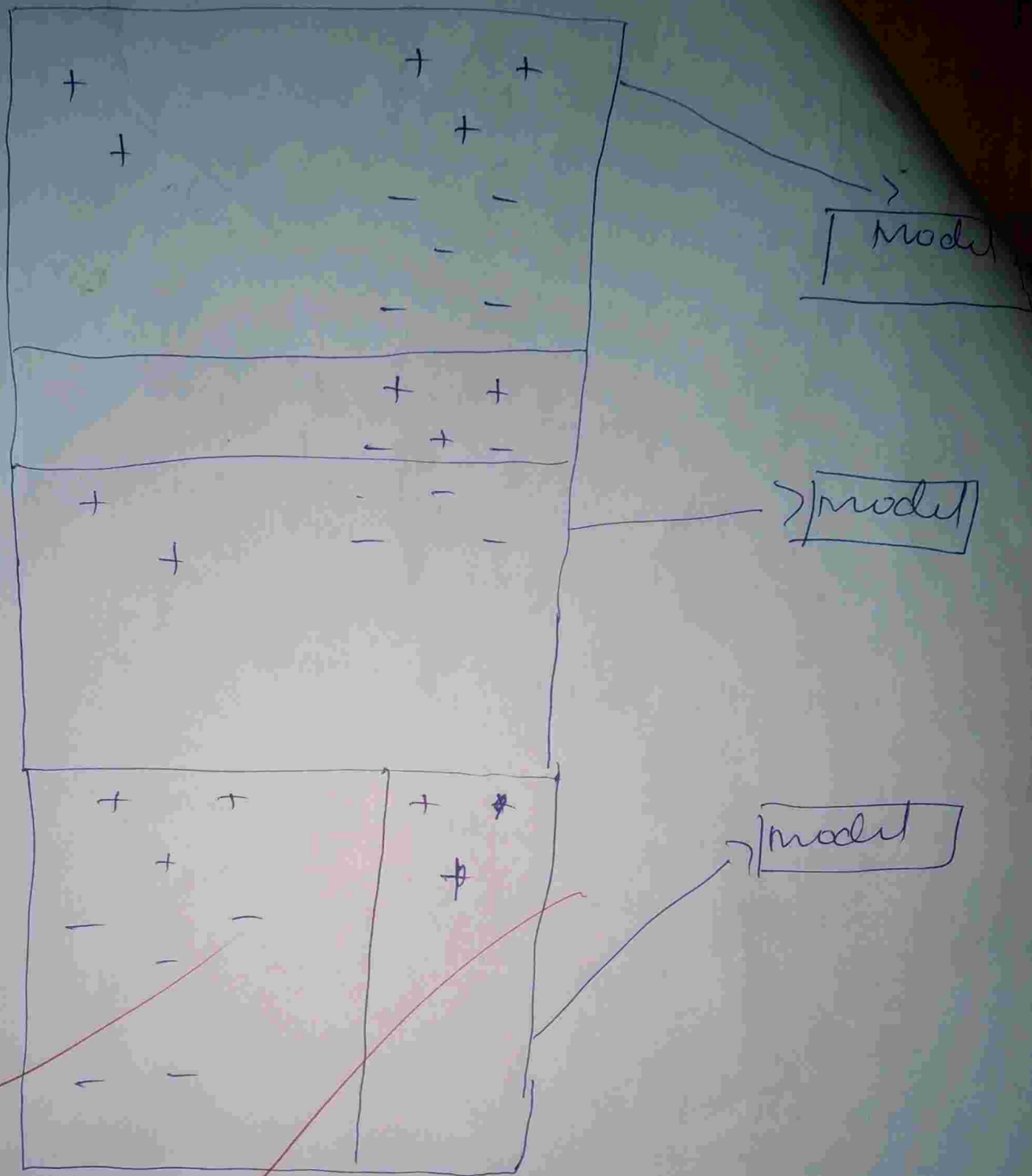
if

* cw to set 5

else

* cw to set 2

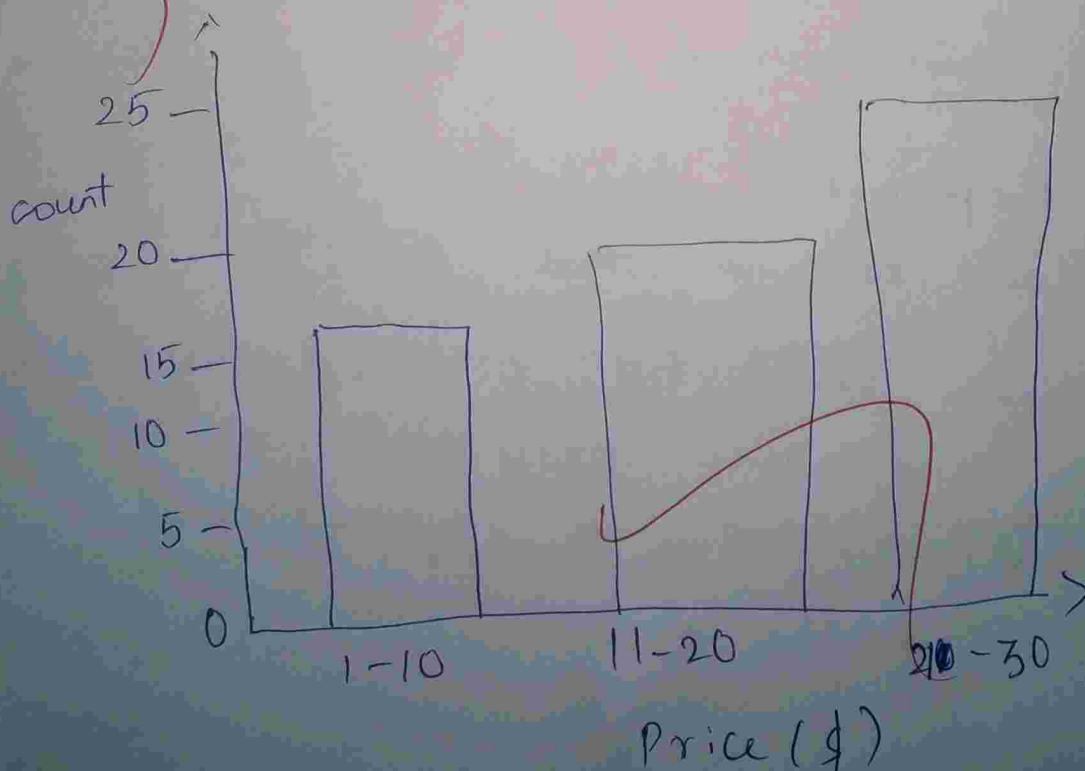
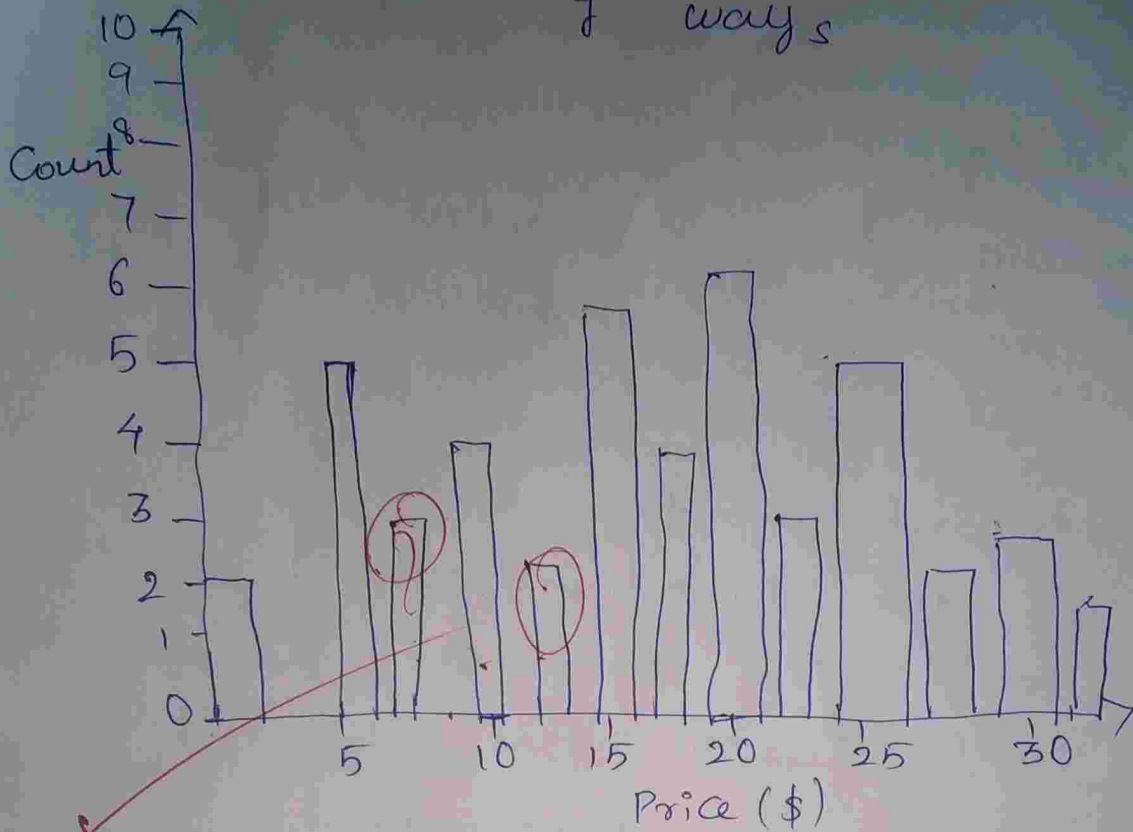
* It is a technique which is used to transform the data in a useful and efficient format.



* These are all the model to achieve this algorithm and to find the original data transformation for that model.

Histogram

The numbers have been sorted in the following ways



* These are the buckets determined in the following attributes and values partitioned.

3.) Data cleaning

* Data cleaning is a removal of noisy data and some unwanted or duplicate data values.

* Insert a record in the table and ~~the~~ set the data in the data values, and the original values are replaced by the general values and these are calculated by a data bins.

* This is a preprocessing technique to insert a data set in the database

* This process has removing or reducing the cache files and the original data has been executing in a efficient manner.

It is the one of the wide techniques the data mining.

* Probably this cleaning process is to remove the unwanted ~~cache files~~ and lossy data.

* It could be updated, mean while the transformation of the data has been occurred in the cleaning process.

* It is used to reduce the number of missing values in the data set.

* The data set is based on the values we have inserted in the table and it would be viewing format and dont need to be change in the record.

* The general data is to be reduce the available procedures to reset the data dependencies.

* These are all the process and to approaches to data cleaning as a process.

4.) k-means partitioning Algorithm

* This algorithm is used to discretization process of the knowledge discovery in database.

* Code patterns is to be executed in this technique and to estimate the predictive method and to identify the number of algorithms the partitioning method, whether the statement is a reputed through the applications. To generate the frequency of the transformation is a more meaningful and which is used to transform the raw data in an useful and efficient format.

* These process to identify those method to acquires of a data output is to be a k-means partitioning algorithm.

5.)

* The simple Bayesian belief network for the variable lung cancer with the following ways

i) Acyclic graph



* This graph represents the determination of increasing the variance from the acyclic graph.

conditional probability table

count	variance	
10		
20		
30		
40		
50		
60	60	
70		
80	80	
90	90	
100		

* This Bayesian belief network represents the increasing the number of count.

* Based on the count and the variance is passed through the particular network variable.

* These are all the network variable in the lung cancer with the following process executed in the table.