Priyanshi Singhal
03815602717
CSE-FICA3

Q// Machine learning Method to predict how people would rate movies, book etc.

Answer

Software used :- Jupyter/python

Algorithm :-

→ Load the data

→ Use KNN algorithm to classify the data

KNN algorithm :-

Let's take a simple case to understand this algorithm following is a spread of red circles (RC) & grey square (GS).

your intend to find out the class of the blue star, BS can either be RL or GS & nothing else.

The K in KNN is the nearest neighbour we wish to take vote from. Let's say K=3, Hence we will now make a circle with BS as center just as big as to enclose only three data points on the plane.

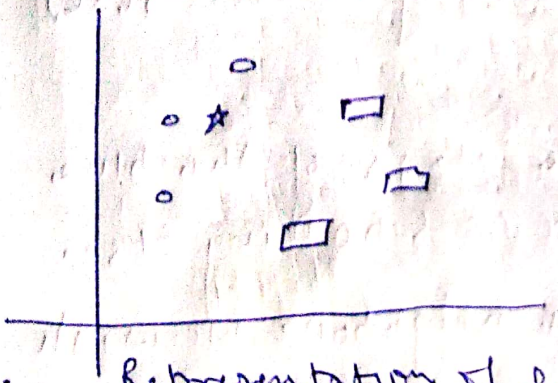
Ref to fig 2

The three closest points BS is all RC. Hence with good class RL there, the choice became very obvious as all three votes from the closest neighbours went to RC. The choice of the

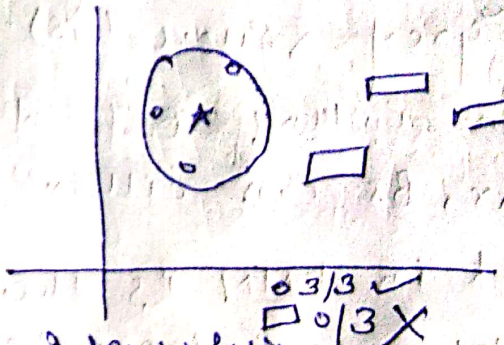
parameters is very crucial in this algo.

For getting the predicted class iterate from 1 to total number of training data points.

1. Cal the distance b/w test data & each row of training data.
   - Here we will use eucledian distance as our distance metric since it's the most popular method. The other metrics that can be used are sine, cosine e.t.c.
2. Sort the calculated distances in ascending order based on distance values.
3. Get top k rows from the sorted array.
4. Get the most frequent class of these rows.
5. Return the predicted class.



Representation of dataset



Identifying closest region

Conclusion:- KNN algo is used as distinct of movie for classification using their ratings so

Q. Estimate the precision, recall, accuracy & f-measure of the decision tree classified.

Answer

Software used: Jupyter / python / WEKA.

Algorithm:-

→ Load the text classification data from web link for dataset.

Apply decision based classifier to calculate precision recall accuracy & F score,

Accuracy :- No of data samples correctly classified / total no. of samples.

F Scores :- The prediction f-measure or balanced f-score is the heremonic mean of precision & recall.

Decision tree classifier :-

→ Algorithm will perform following task recursively:

1. create root node for the tree.
2. If all examples are positive, return leaf node 'positive'
3. Else if all examples are negative, return leaf node 'negative'
4. calculate the entropy of current state.
5. For each attribute, calculate the entropy with respect to the attribute denoted by H(S)
6. Select the attribute which has more value of IG
7. Remove the attribute that offers highest IG from set of attrib

Conclusion:-

Algorithm has been run on all the programs for given dataset