ID3 Algorithms in Decision Trees

Soumya Dasgupta, Akankshya Nayak

National Institute of Science Education and Research

February 13, 2023

CONTENTS

Part I

ID3 ALGORITHM

UNDERSTANDING DECISION TREES

- A decision tree is a tree in which a decision is taken at every node. The leaf nodes of the tree generally indicate the final decision of the tree. The set of questions that are asked to take a decision are known as features. Through the answers to these features, the decision tree reaches at a conclusion usually termed as the label.
- There are multiple algorithms to create decision trees. One such algorithm is ID3.

INTRODUCTION TO ID3

- ID3 stands for Iterative Dichotomiser 3 which was first invented by Ross Quinlan
- Iteratively (repeatedly) Dichotomizes (divides) the features into groups
- Top-down (builds the tree from the top), greedy (at each step, selects the current best feature to create a node) approach to build decision trees
- Classification of nominal data

ID3 ALGORITHM

- ▶ Finds the best feature on the basis of Information Gain or Gain
- ▶ Information Gain tries to minimize the entropy in the data set i.e. the measure of disorder in the target feature. Entropy of a dataset S is denoted as:

$$Entropy(S) = -\Sigma p_i * \log_2(p_i); \ i = 1 \text{ to } n$$

  Where, n: Total no. of classes in target column, p_i: Probability of class i in the target column

ID3 Algorithm

▶ Then, Information Gain of a particular feature column A of the dataset S is calculated as:

$$IG(S, A) = Entropy(S) - \Sigma((|S_V|/|S|) * Entropy(S_V))$$

Where, S_V: Set of rows in S for which the feature column A has value V, $|S_V|$: Number of rows in S_V, $|S|$: Number of rows in S **r2**

INDUCTIVE BIAS IN ID3 ALGORITHM

- Inductive bias is based on the ordering of hypotheses by search strategy
- Approximate inductive bias of ID3: Shorter trees are preferred over Larger trees.
- Trees that place high information gain attributes close to the root are preferred over those that do not.

MAJOR STEPS

- Calculation of Information Gain(IG)
- If all rows don't belong to the same class, find the feature with maximum Information Gain and split the dataset into subsets on the basis of this feature
- Make a decision tree node using this feature
- If all the rows belong to the same class, the current node is assigned a leaf node with a label of that class
- Repeat for every feature
- Terminate when all the features are over or the decision tree contains only leaf nodes
- The following website can be visited to understand ID3 with an example: `https://sefiks.com/2017/11/20/a-step-by-step-id3-decision-tree-example/`

ADVANTAGES

- ▶ Prediction rules are created from the training data and are easily understandable
- ▶ Creates a short tree in relatively less time
- ▶ It only needs to test enough attributes until all data is classified
- ▶ Finding leaf nodes enables test data to be pruned, reducing the number of tests

DISADVANTAGES

- ▶ Data may be over-fitted or over-classified, if a small sample is tested.
- ▶ Only one attribute at a time is tested for making a decision.

REFERENCES

- Machine learning/inductive inference/decision trees/construction. Available at:
  `www.cs.ucdavis.edu/~vemuri/classes/ecs271/Decision%`
  `20Trees-Construction.htm#:~:text=Approximate%20inductive%20bias%20of%`
  `20ID3,over%20those%20that%20do%20not.https://`
- Sakkaf, Y. (2020) Decision trees for classification: Id3 Algorithm explained, Medium. Towards Data Science. Available at: `https://towardsdatascience.com/`
  `decision-trees-for-classification-id3-algorithm-explained-89df76e72df1`
- Mantri, N. (2021) Using ID3 algorithm to build a decision tree to predict the weather, OpenGenus IQ: Computing Expertise amp; Legacy. Available at:
  `https://iq.opengenus.org/id3-algorithm/`
- Serengil, S. (2021) A step by step id3 decision tree example, Sefik Ilkin Serengil. Available at:
  `https://docs.rapidminer.com/latest/studio/operators/modeling/`
  `predictive/trees/id3.html`
- GmbH, R.M. Id3 (RapidMiner Studio Core), ID3 - RapidMiner Documentation. Available at:
  `https:`
  `//sefiks.com/2017/11/20/a-step-by-step-id3-decision-tree-example/`