ML Notes

0.1 Decision Trees

Source

- https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

Keywords

- supervised learning

0.1.1 ID3

The ID3 algorithm builds decision trees using a top-down greedy search approach through the space of possible branches with no backtracking.

1. It begins with the original set S as the root node.
2. On each iteration of the algorithm, it iterates through the very unused attribute of the set S and calculates **Entropy(H)** and **Information gain(IG)** of this attribute.
3. It then selects the attribute which has the smallest Entropy Largest Information gain.
4. The set S is then split by the selected attribute to produce a subset of the data.
5. The algorithm continues to recur on each subset, considering only attributes never selected before.

Attribute Selection Method

If the dataset consists of **N** attributes then deciding which attribute to place at the root or at different levels of the tree as internal nodes is a complicated step. By just randomly selecting any node to be the root can't solve the issue. If we follow a random approach, it may give us bad results with low accuracy.

For solving this attribute selection problem, researchers worked and devised some solutions. They suggested using some *criteria* like :

Entropy,
Information gain,
Gini index,
Gain Ratio,
Reduction in Variance
Chi-Square

If we take entropy as selection, ID3 follows the rule — A branch with an entropy of zero is a leaf node and A brach with entropy more than zero needs further splitting.

0.2 Differential Privacy

- https://www.youtube.com/watch?v=MOcTGM_UteM

1 Theos Thesis