



ML Day2, 3

Algorithms

ID3 Algorithm(Iterative Dichotomizer 3)

- the basic approach to decision tree induction
- 반복적으로 이분하는 알고리즘(root node에서 leaf node를 찾을 때까지 계속 반복적으로 계산과정이 진행된다)
- 주어진 데이터를 이용해 the shallowest decision tree를 만드는 알고리즘
- top-down(불순도를 낮추는 방향으로)방식
- recursive(반복, 계속적으로)
- depth-first partitioning
- entropy를 불순도로 계산, 독립변수가 모두 범주형일때만 가능하다.
 - ID3 Algorithm은 categorical features and clean data에 자주 쓰이는 경향이 있다.

- Information Gain의 한계점: IG의 수치가 높아 겉보기엔 잘 분류됐다고 보이지만 구체적으로 따져봤을 때 불필요하게 많은 class로 분류가 될 가능성이 있다.