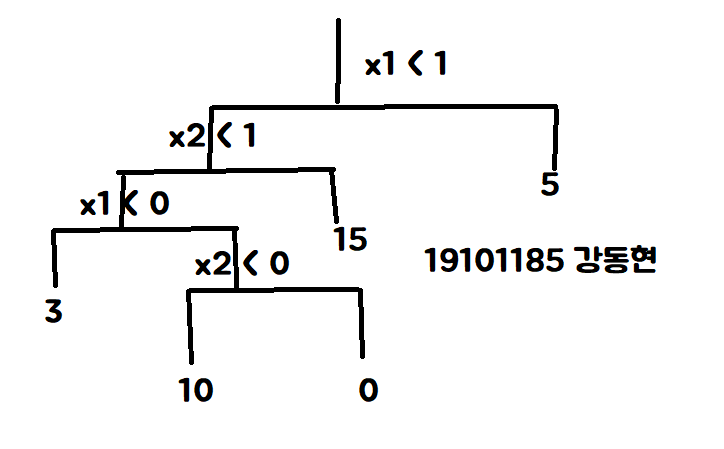
19101185 강동현 데이터분석 Assignment 10

3. 지니 인덱스와 분류 error , 엔트로피를 2개의 클래스의 simple classification을 본다. Single plot을 각 경우에 따라 보여야 한다. 지니 인덱스와 엔트로피 , 분류 에러에 대해서 식을 정리하면 다음과 같다. 지니는 시그마 p(1-p) 로 표현된다. 이 경우 지니는 p에 대한 2차함수 형태로 표현된다. 또한 p = 1-p일 때 최대값을 가지므로 p = 1/2일 때 최대값을 가지는 이차함수의 형태로 표현이 될 것이다. 엔트로피의 경우는 -시그마pLogp 이다. 힌트에 1-p도 이용하라고 했으므로 본 식에 1-p를 대입해서 두 식을 더해준 값이 엔트로피의 경우이다. 따라서 -시그마plogp + -시그마(1-p)log(1-p) 이다. 엔트로피 역시 1-p와 p사이의 관계가 존재하므로 p = 1-p 인 p = 1/2에서 최대값을 가질 것이다. E 즉 Classification Error는 1 – maxk(p) 이다. 이 경우 역시 문제에서 2개의 클래스만 고려하므로 maxk 부분에서 p와 1-p 두가지만 고려하면 된다. 즉 이 경우 나올 수 있는 1-p 혹은 p값의 최대값은 1/2가 될 것이다. E는 다른 그래프와는 다르게 1 – maxk(p) 즉 일차식의 모양이기 떄문에 p = 1/2에서 최대값을 가지지만 지니 인덱스와는 다르게 일차식의 모양으로 각지게 그려질 것이다.

4. 그림 8.12를 보고 답하는 문제이다.

(a). 그림 8.12의 왼쪽 사진을 보고 tree 형태로 표현하는 문제이다.



(b). 8.12 그림 오른쪽 트리를 보고 구역을 나누어 다이어그램을 그리는 문제이다.

