ML Day4

- **Continuous Descriptive Feature**
  - continuous feature values sorting(연속적인 feature를 오름 또는 내림차순으로 정렬)
  - target feature level이 다른 instance들이 인접해 있을 때(다른 클래스로 변하는 지점) 그 사이를 threshold(분기 지점) 후보로 체크
  - threshold = 변하는 지점에 인접한 클래스들의 평균 값
  - I.G값이 가장 높은 threshold를 선택 (Determining the BEST Threshold!!)
  - continuous feature는 한 path상에서 여러 번 이용될 수 있다.

- **Continuous Target Feature**
  - Regression Tree
  - value output by the leaf node

    ⇒ leaf node 안에는 instance들의 mean(평균)값이 표기됨
  - leaf node 안 target feature 값들의 variance가 감소하는 방향으로(비슷한 target feature value를 가진 instance들끼리 뭉치도록 해줌)

⇒ best feature를 선택함에 있어서, ID3에서 entropy를 이용했듯이 variance값을 이용

- variance(분산 = 편차 제곱의 평균)

  ⇒ 편차: 원래의 값에서 평균을 뺀 값

  ⇒ 표준편차: 평균 값이 실제 값에서 부터 얼마나의 오류가 있느냐

  ⇒ 분산: 편차들을 합하기 전에 제곱을 해서 합한 것(떨어진 정도를 파악하는데 음수와 양수를 가진 값이 서로 상쇄하여 제대로 판별할 수 없는 경우가 있으므로)