**2.3. Introduction to Decision Tree**

**DECISION TREE**

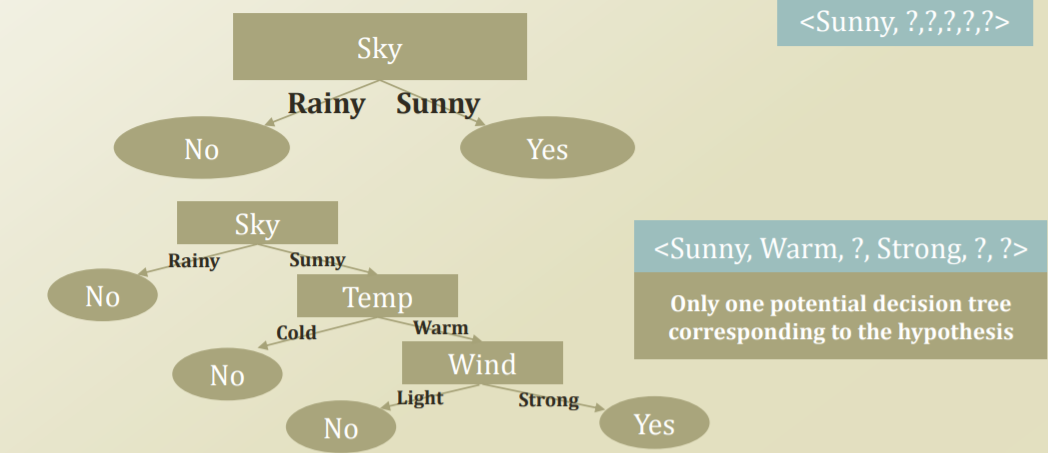


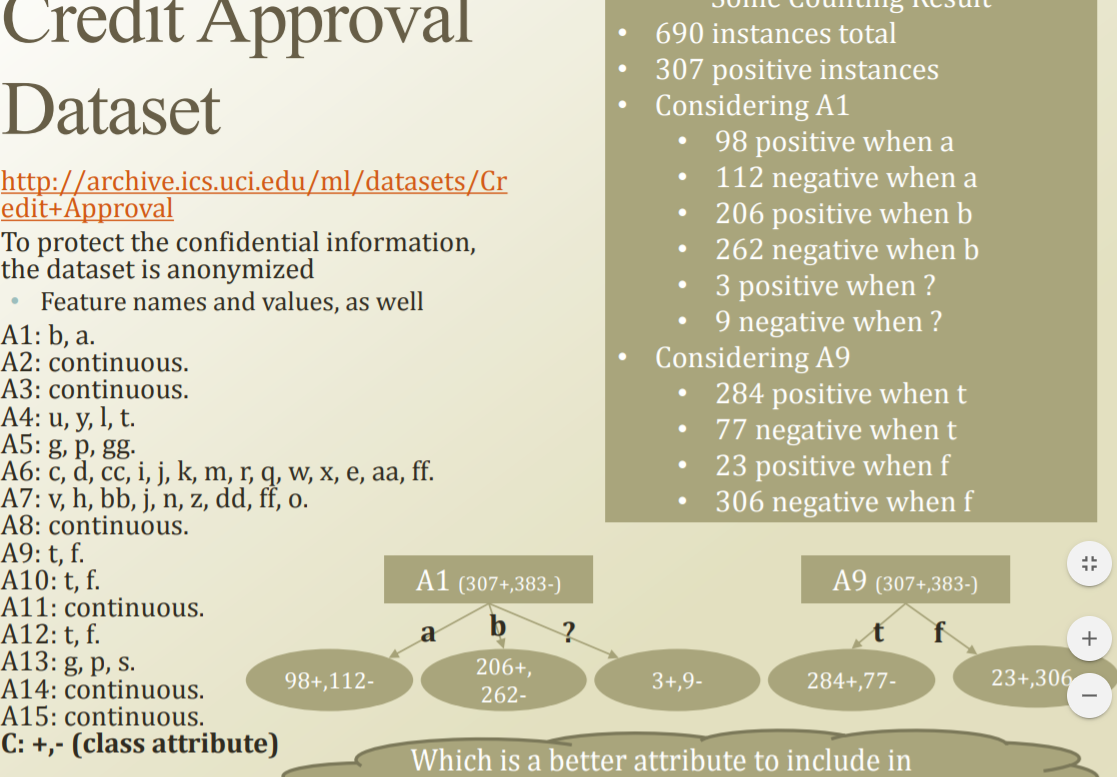
We need a better learning method

• We need to have more robust methods given the noises

• We need to have more concise presentations of the hypotheses

• One alternative is a decision tree



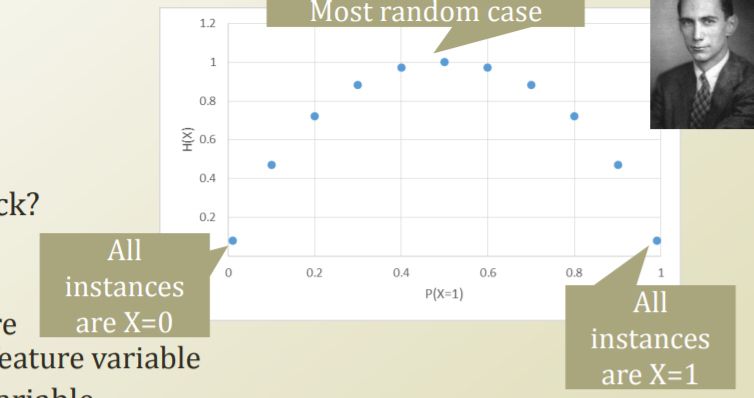


Which is a better attribute to include in the feature set of the hypothesis?

A1~15 case가 있으며 +- 는 그 칼럼의 나가서논다, 안나가서 논다. 2가지 case입니다.

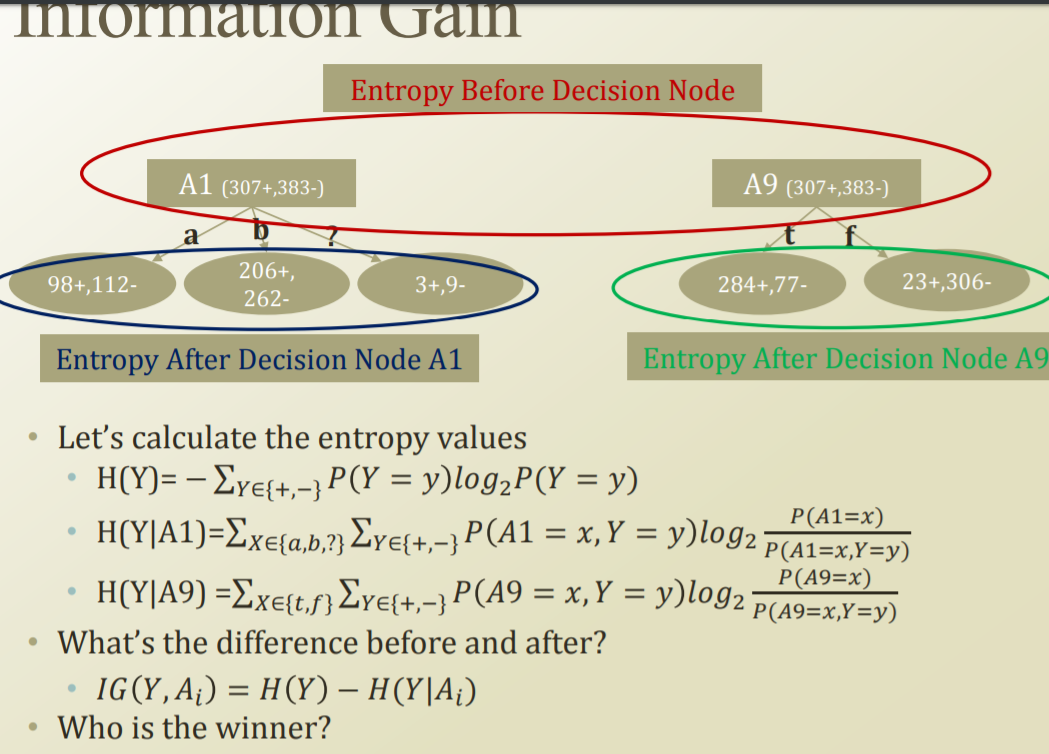
A1 > a b ? ?는 기타 데이터 A1중 a는 98은 나가고 112는 안나간다.

**Entropy**

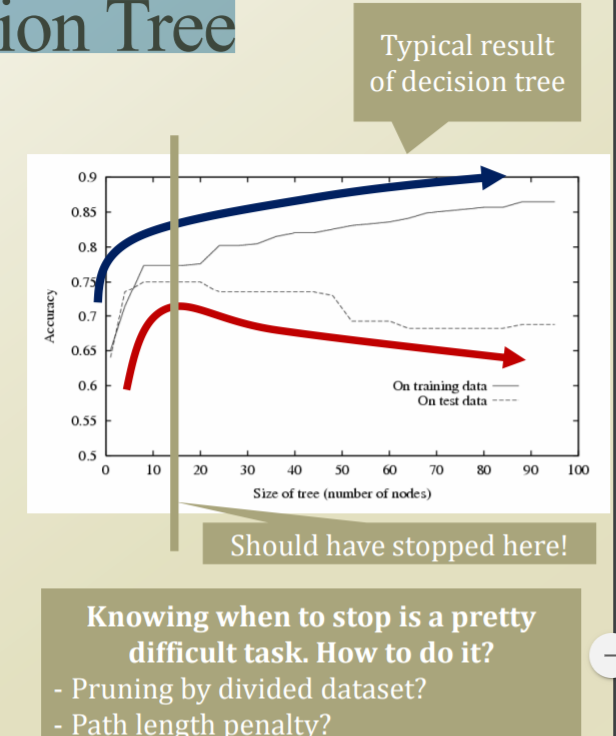


• Better attribute to check? • Reducing the most uncertainty • Then, how to measure the uncertainty of a feature variable • Entropy of a random variable • Features are random variables • Higher entropy means more uncertainty • 𝐻 𝑋 = − σ𝑋 𝑃 𝑋 = 𝑥 𝑙𝑜𝑔𝑏𝑃(𝑋 = 𝑥) • Conditional Entropy • We are interested in the entropy of the class given a feature variable • Need to introduce a given condition in the entropy • 𝐻 𝑌|𝑋 = σ𝑋 𝑃 𝑋 = 𝑥 𝐻 𝑌 𝑋 = 𝑥 = ෍𝑋 𝑃 𝑋 = 𝑥 {−෍𝑌 𝑃 𝑌 = 𝑦|𝑋 = 𝑥 𝑙𝑜𝑔𝑏𝑃 𝑌 = 𝑦 𝑋 = 𝑥 }

Information gain



Problem of Decision Tree

 • We did better in the given dataset! • Only in the given experience, a.k.a. Training dataset • What if we deploy the created decision tree in the field? • World has so much noise and inconsistencies. • The training dataset will not be a perfect sample of the real world • Noise • Inconsistencies