Fundamentals of Machine Learning

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Learn the most classical methods of machine learning
  - Rule based approach
  - Classical statistics approach
  - Information theory appraoch
- Rule based machine learning
  - How to find the specialized and the generalized rules
  - Why the rules are easily broken
- Decision Tree
  - How to create a decision tree given a training dataset
  - Why the tree becomes a weak learner with a new dataset
- Linear Regression
  - How to infer a parameter set from a training dataset
  - Why the feature engineering has its limit

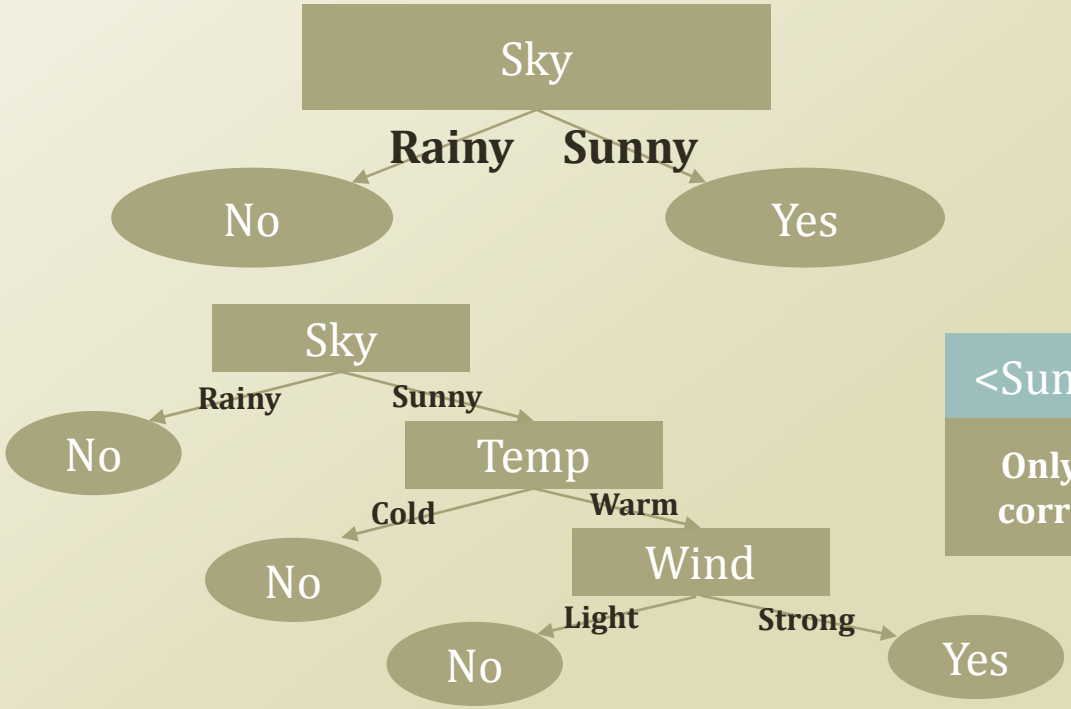
DECISION TREE

Because we live with noises…

Sky	Temp	Humid	Wind	Water	Forecst	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

- 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

<Sunny, ?,?,?,?,?>



<Sunny, Warm, ?, Strong, ?, ?>

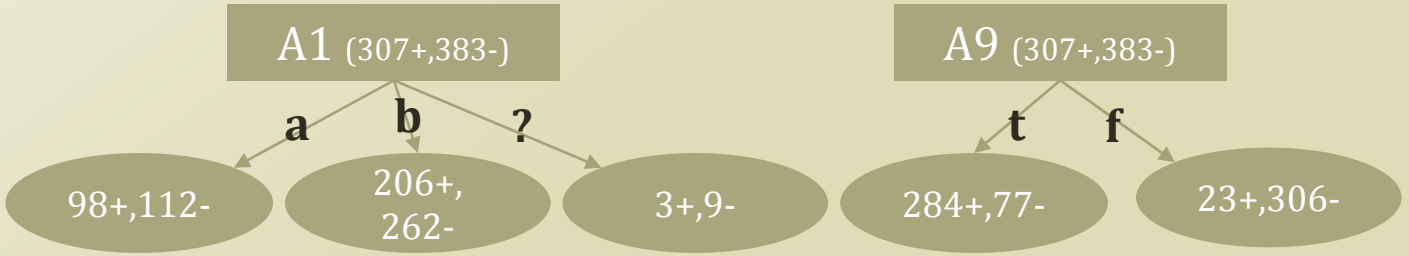
Only one potential decision tree corresponding to the hypothesis

Credit Approval Dataset

- http://archive.ics.uci.edu/ml/datasets/Credit+Approval
- To protect the confidential information, the dataset is anonymized
  - Feature names and values, as well
- A1: b, a.
  A2: continuous.
  A3: continuous.
  A4: u, y, l, t.
  A5: g, p, gg.
  A6: c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.
  A7: v, h, bb, j, n, z, dd, ff, o.
  A8: continuous.
  A9: t, f.
  A10: t, f.
  A11: continuous.
  A12: t, f.
  A13: g, p, s.
  A14: continuous.
  A15: continuous.
  C: +,- (class attribute)

Some Counting Result

- 690 instances total
- 307 positive instances
- Considering A1
  - 98 positive when a
  - 112 negative when a
  - 206 positive when b
  - 262 negative when b
  - 3 positive when ?
  - 9 negative when ?
- Considering A9
  - 284 positive when t
  - 77 negative when t
  - 23 positive when f
  - 306 negative when f



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