11/26/13 Recitation

Naive Bayes and Bayesian Networks

Naive Bayes Background

- Classification algorithm
- Requires discrete features
- Popular, fast, easy to implement
- Versatile: text classification, weather, computer vision
- Performs well even when main assumption is violated

Naive Bayes Details

- Classification rule: choose most likely label given features:
  - Pick y to max P(y | x) = P(x | y) P(y) / P(x) [Bayes]
  - Equivalent to maximizing P(y, x) = P(x | y) P(y)
- Assume conditional independence of features x
  - P(x1, x2, x3 | y) = P(x1 | y) P(x2 | y) P(x3 | y)
  - If no conditional independence, given d binary features and binary y, what is the order of the number of parameters needed to know P(x | y) for all possible x and y?
  - Order of the number of parameters with conditional independence?

Prior and Likelihood

- Final classification rule: max P(y) Prod[P(xi | y)]
- Prior P(y) = (# y) / N
  - Fraction of examples with label y
- Likelihood P(xi | y) = (# y and xi) / (# y)
  - Among examples with label y, the fraction whose i-th feature equals the value xi

Naive Bayes Example

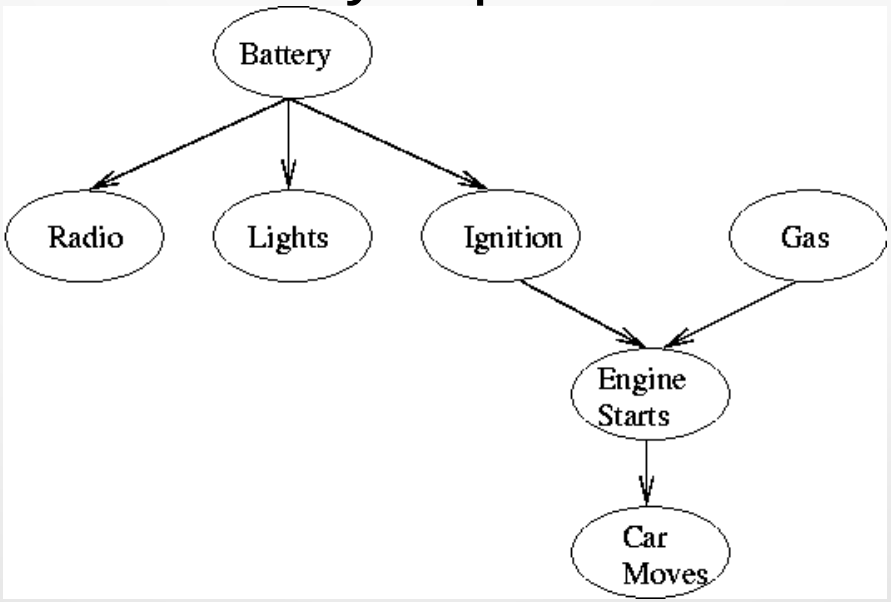
- Label y: does the patient have lung cancer?
  - 1 = yes, 0 = no
- Features x1, x2:
  - x1 whether the patient smokes
  - x2 whether any family member has had lung cancer
  - Are they conditionally independent realistically?

NB Example

- 3 training examples
  - 1. has cancer, smokes, family does not have cancer
  - 2. has cancer, does not smoke, family has cancer
  - 3. no cancer, does not smoke, family does not have cancer
- Classify the following:
  - 4. smokes, family has cancer
  - 5. does not smoke, family has cancer
- Note “smoothing” issue

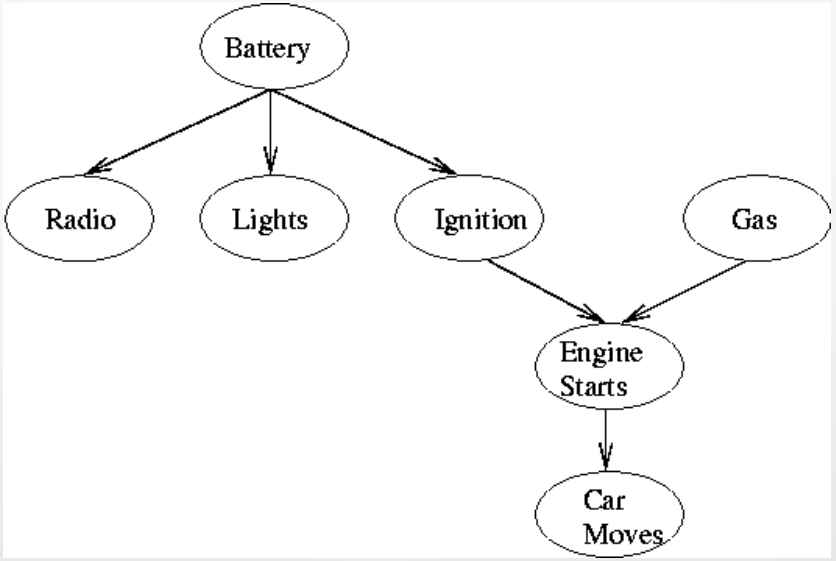
Bayesian Network Background

- AKA Bayes Net
- Graphical model on a directed acyclic graph
- Generalizes Naive Bayes and logistic regression
- Fast
- For inferring the most likely explanation of an outcome



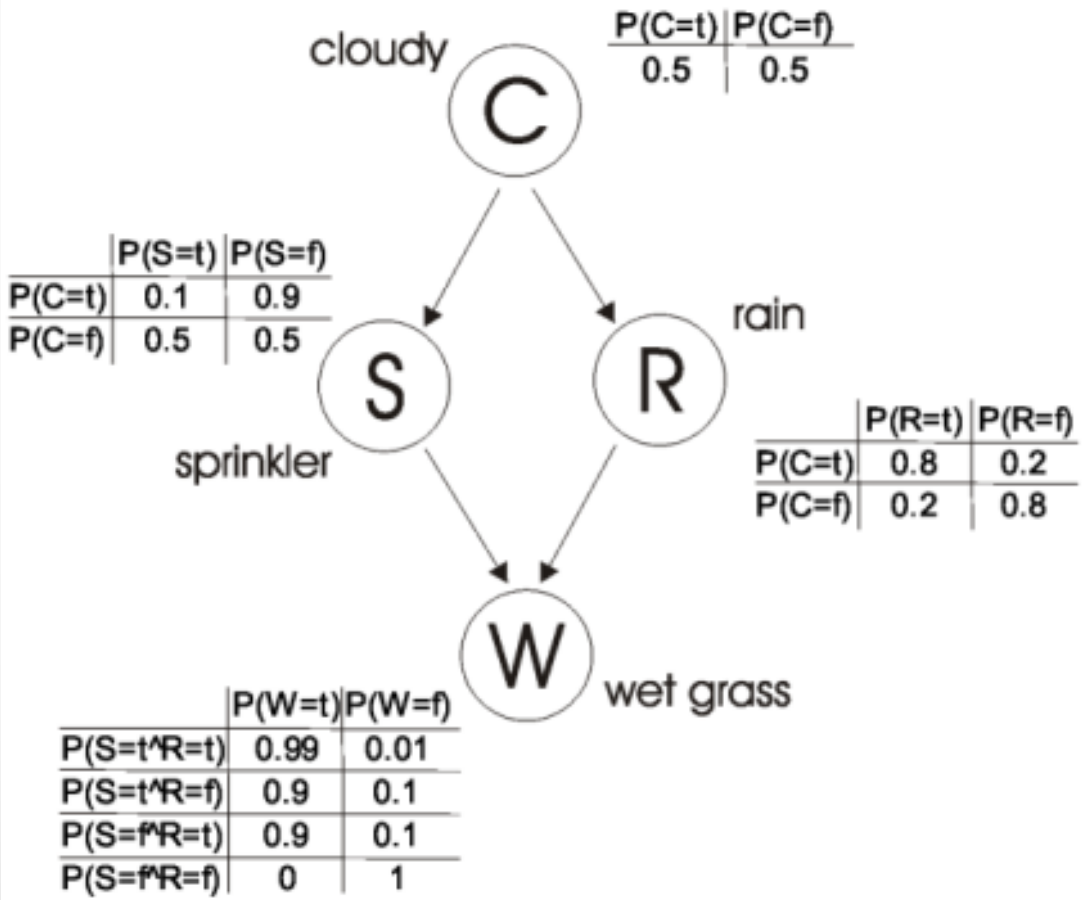
Bayes Net Details

- Nodes = variables
- Edges = dependencies
- Conditional independence assumption: Given all parents of node v, v is independent of non descendants
- Parameters are the conditional probability tables (CPTs) of a variable given all combinations of its parents

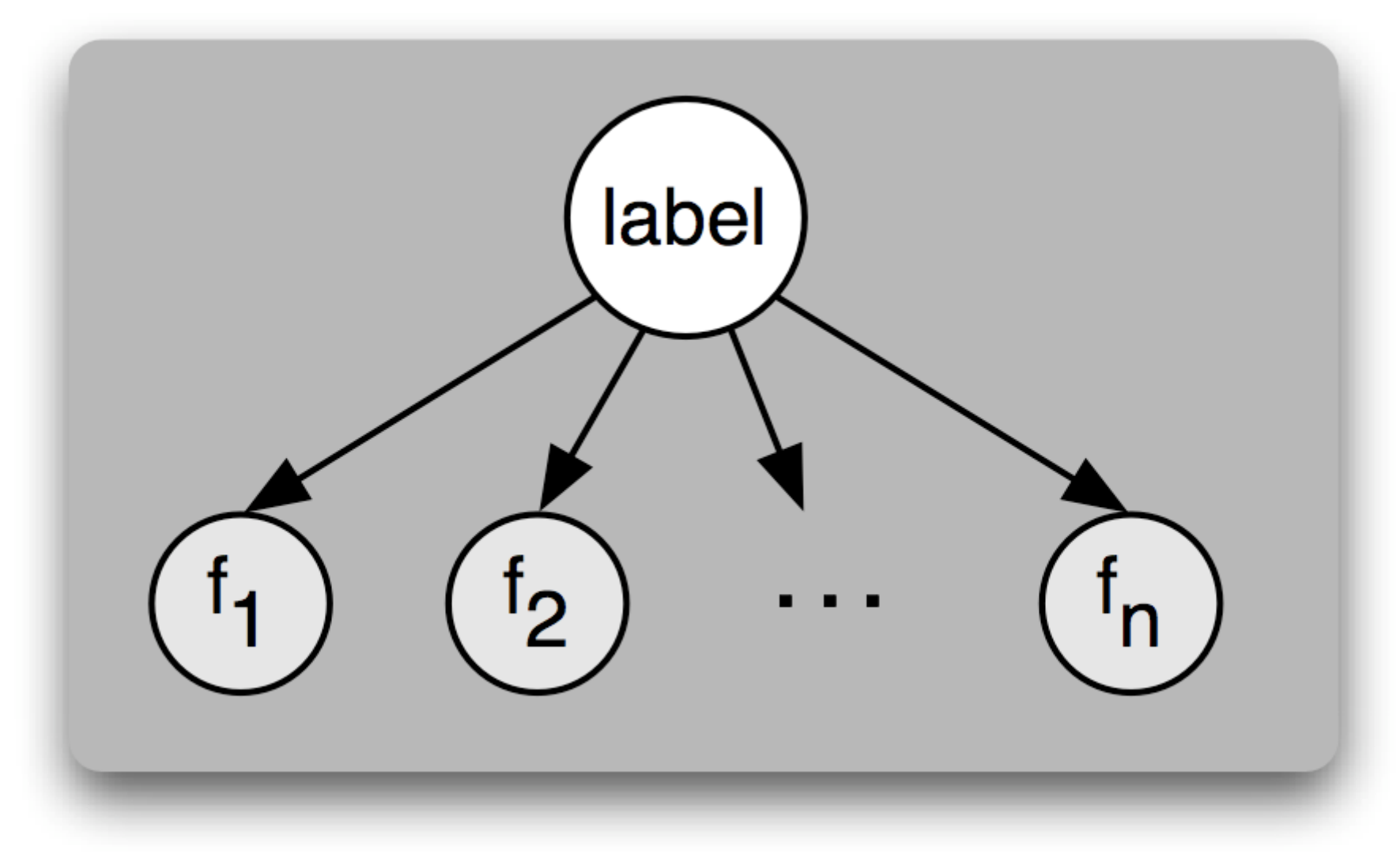


Bayes Net Example

- Joint: P(C,S,R,W) = P(C) P(R|C) P(S|C) P(W|R,S)
- P(C=t | W=t, R=f, S=t) = ?
- P(C=t | W=t, R=f) = ?
  - Inferring variable from evidence
  - Could also infer most likely situation



Naive Bayes as Bayes Net



Efficiency of Bayes Net

- Assume binary variables
- For a node with k parents, what is the order of the number of parameters for its CPT?
- If there are n nodes with at most k parents, what is the order of the total number of parameters?
- If there was no conditional independence, what would be the order of the total number of parameters?