Probabilistic Graphical Models and Bayesian Networks

Machine Learning
CSx824/ECEx242
Bert Huang
Virginia Tech

Outline

- Probabilistic graphical models
- Bayesian networks
- Naive Bayes and Logistic Regression as Bayes nets
- Inference in Bayes nets

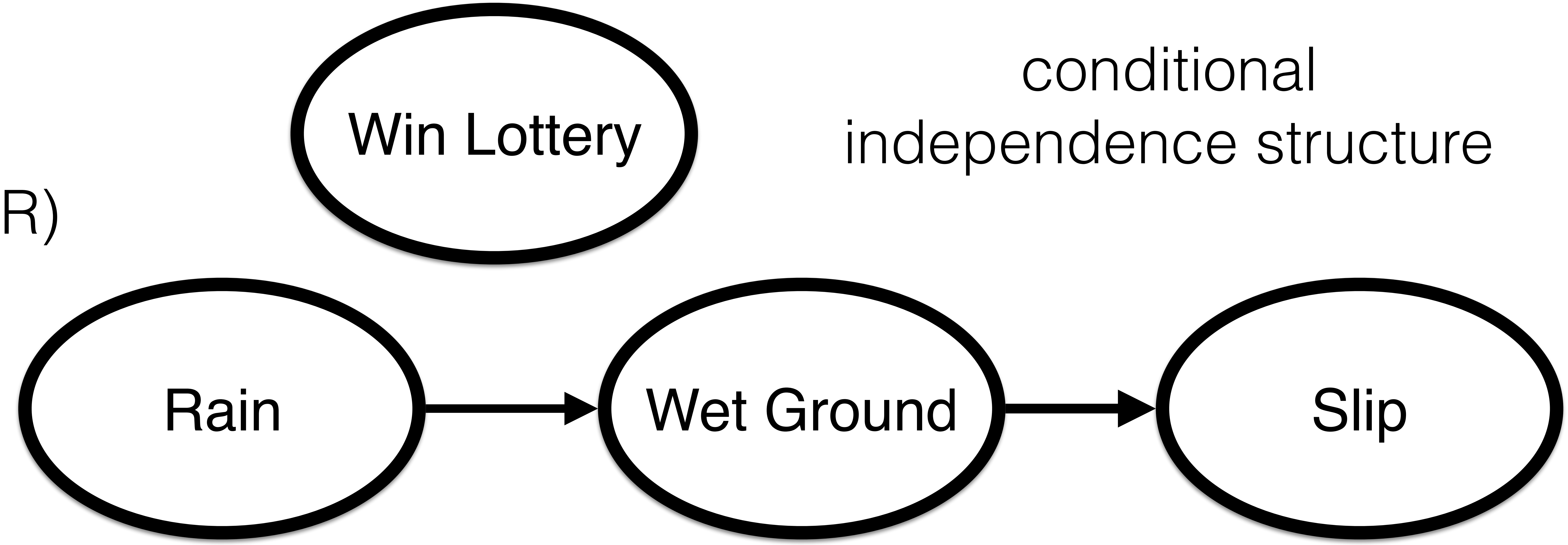
Probabilistic Graphical Models

- PGMs represent probability distributions
- They encode conditional independence structure with graphs
- They enable graph algorithms for inference and learning

Bayesian Networks

P(L, R, W)

= P(L) P(R) P(W | R)



conditional independence structure

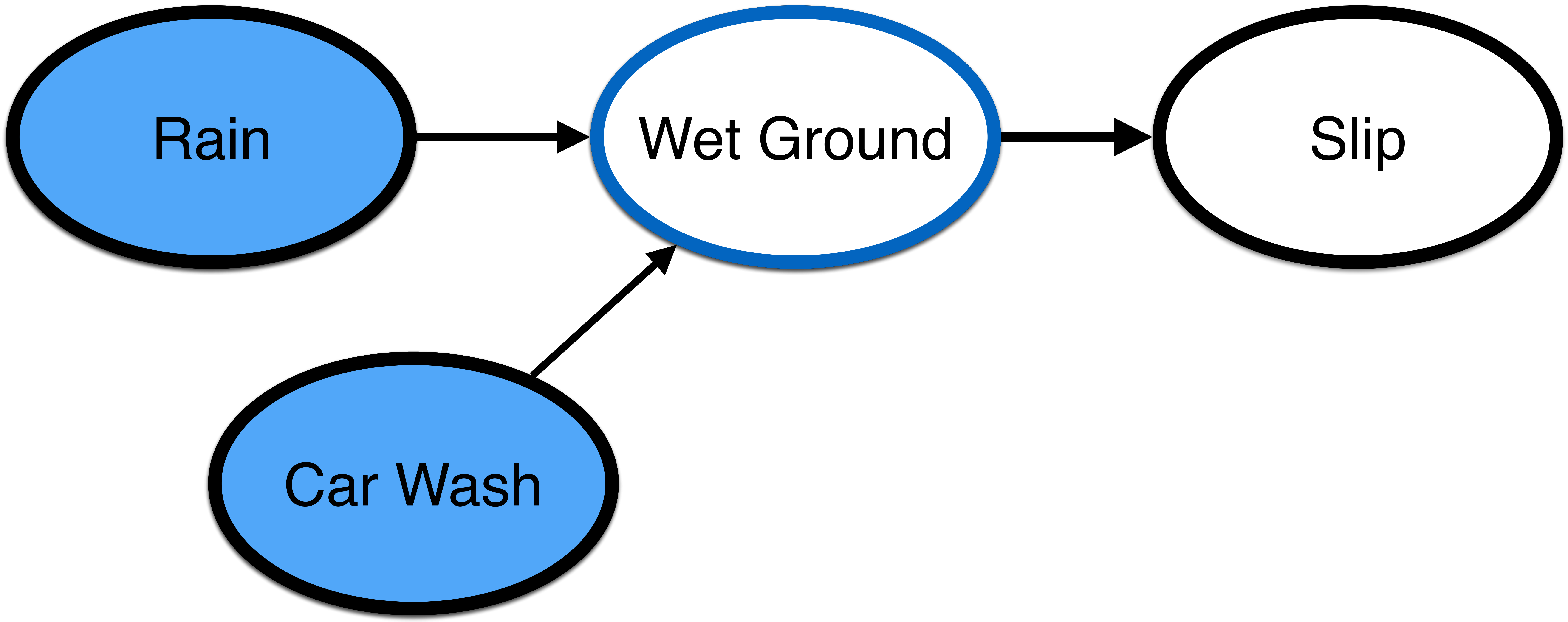
P(L, R, W, S) = P(L) P(R) P(W | R) P(S | W)

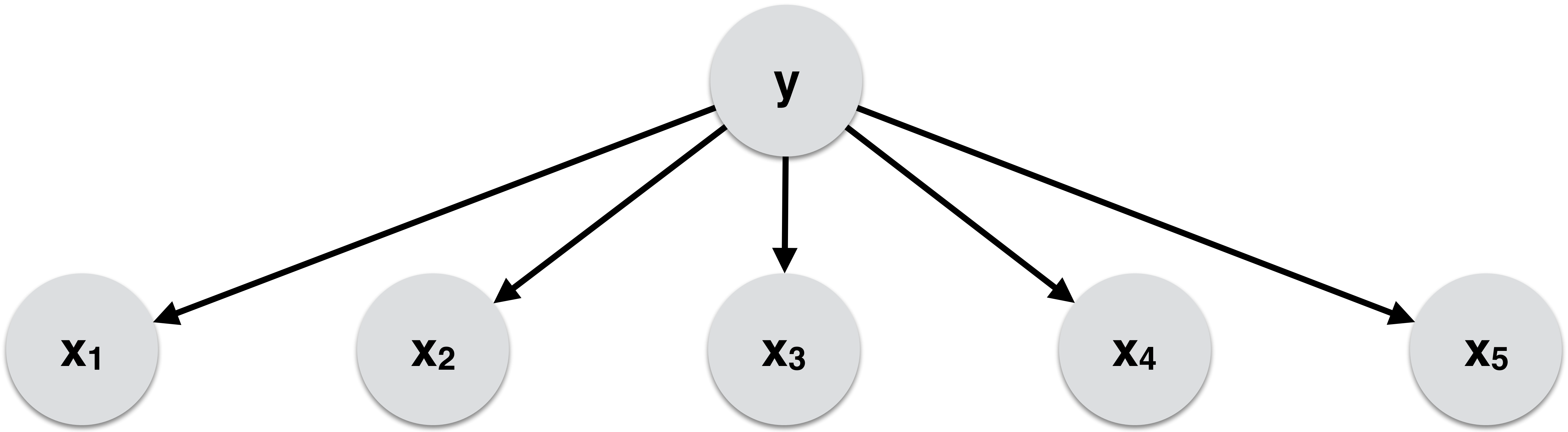
~~P(S | W, R)~~

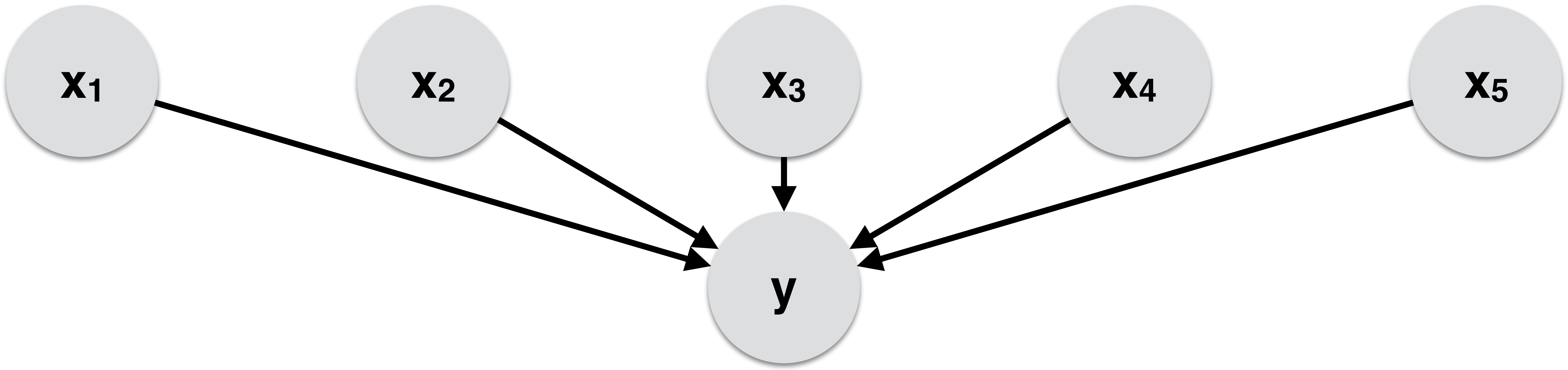
Bayesian Networks

$P(R, W, S, C) = P(R)\, P(C)\, \boxed{P(W \mid C, R)}\, P(S \mid W)$

$P(X \mid Parents(X))$

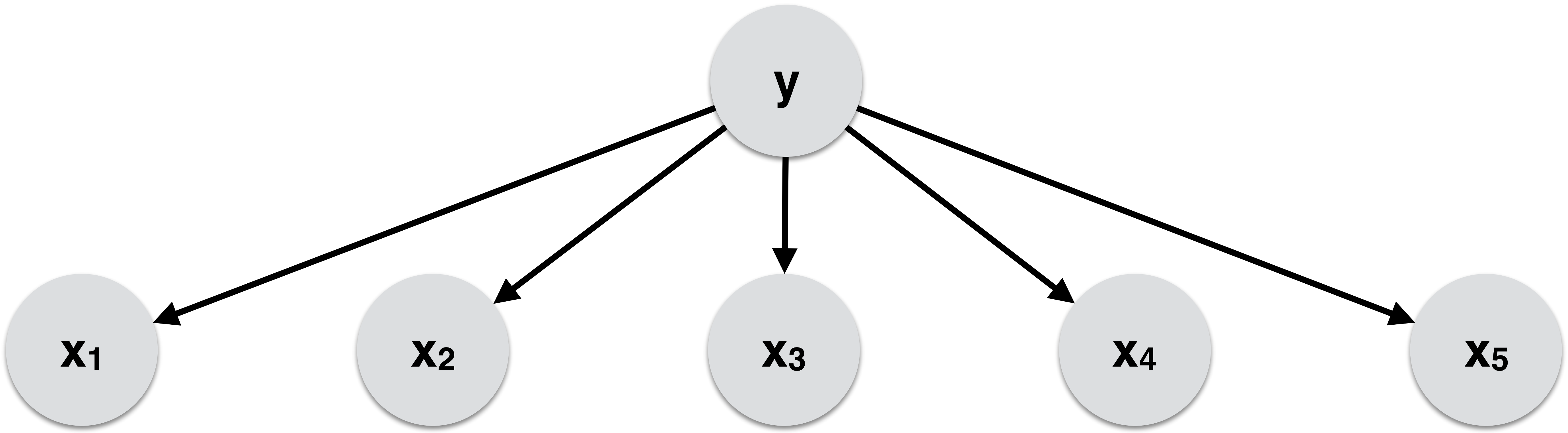


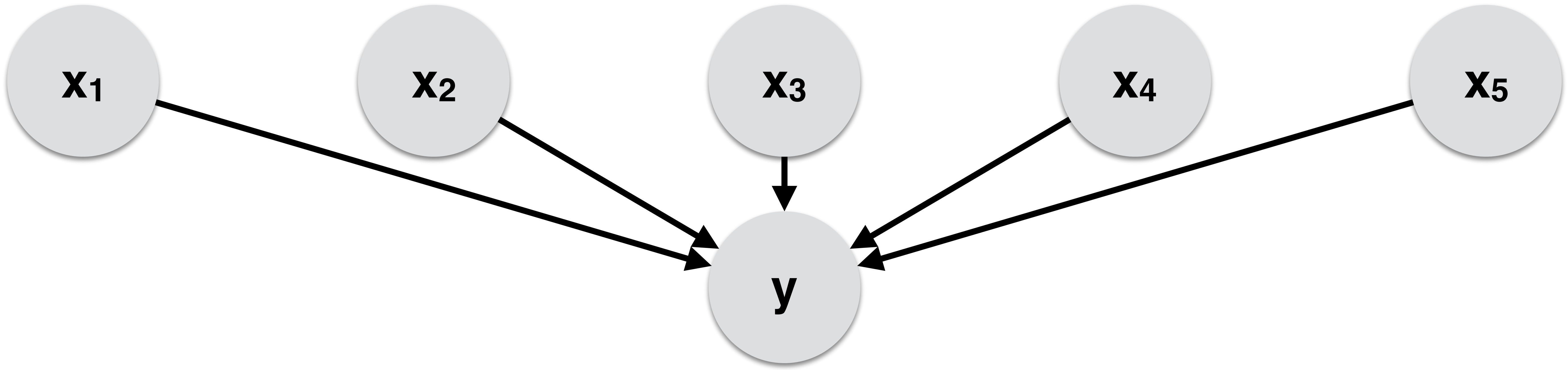

y
x1
x2
x3
x4
x5


x1
x2
x3
x4
x5
y

naive Bayes

$$p(y)\prod_{i=1}^{5} p(x_i|y)$$



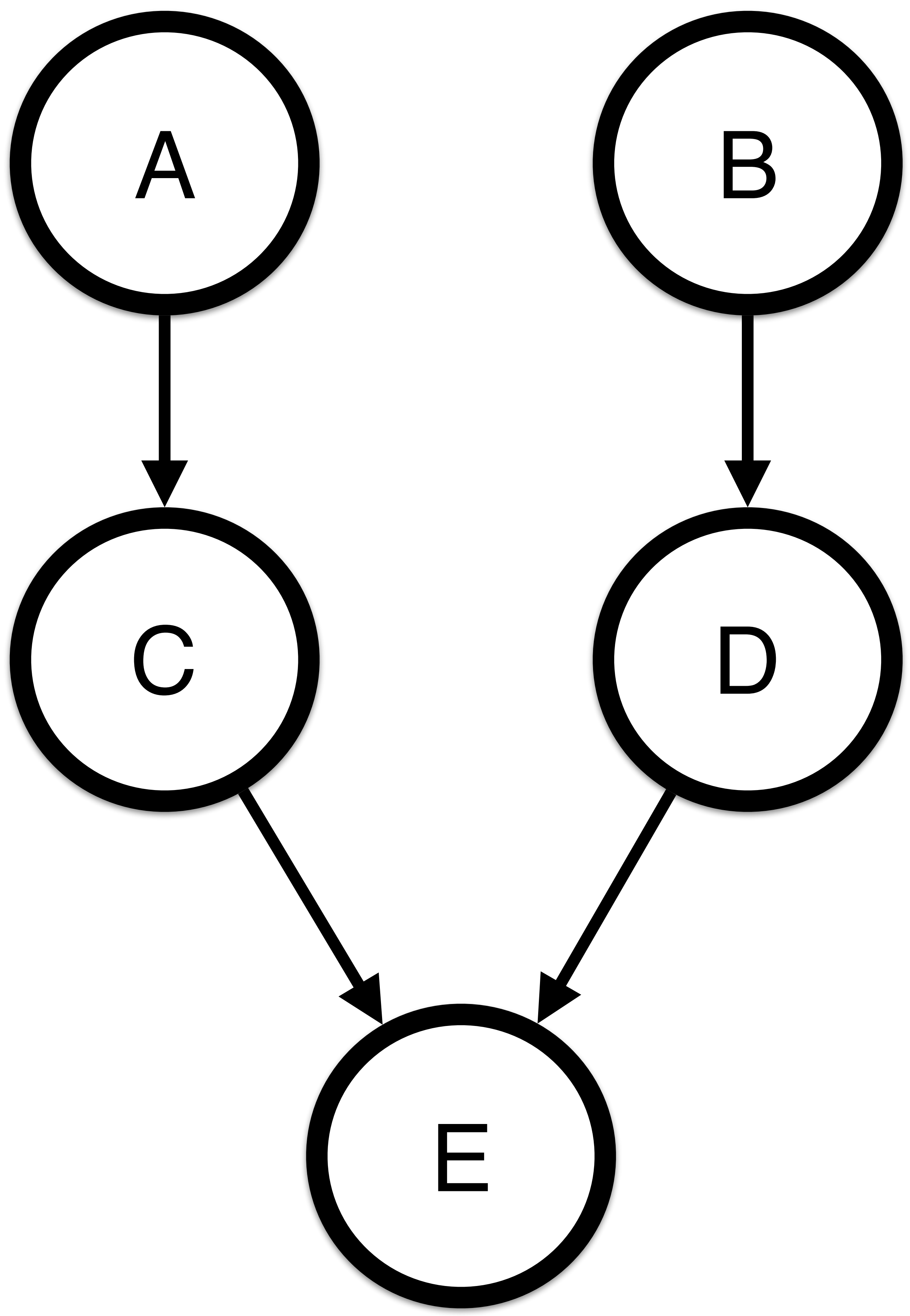


$$p(y|x_1, x_2, x_3, x_4, x_5)\prod_{i=1}^{5} p(x_i)$$

logistic regression (with input likelihood)

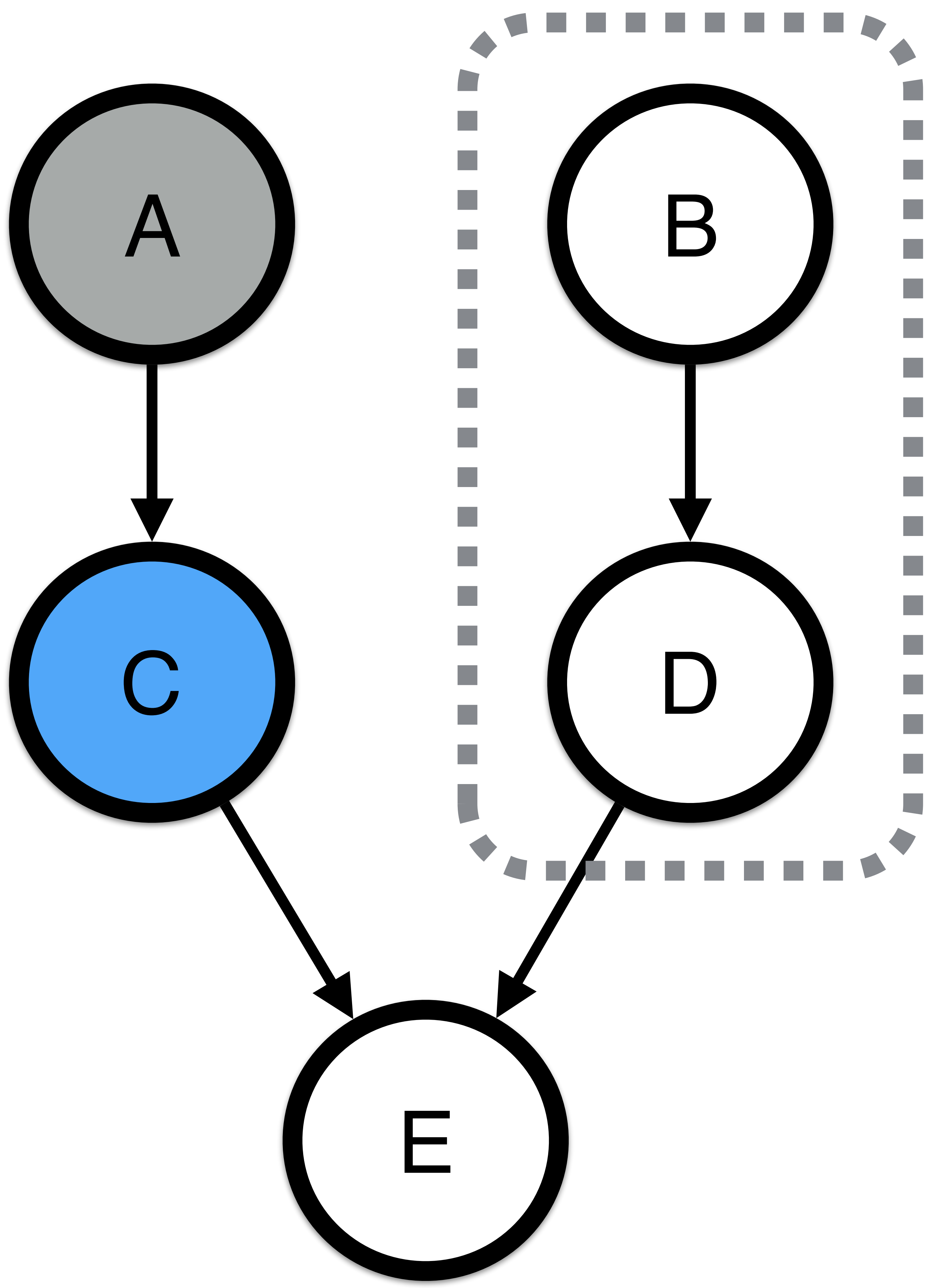
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**



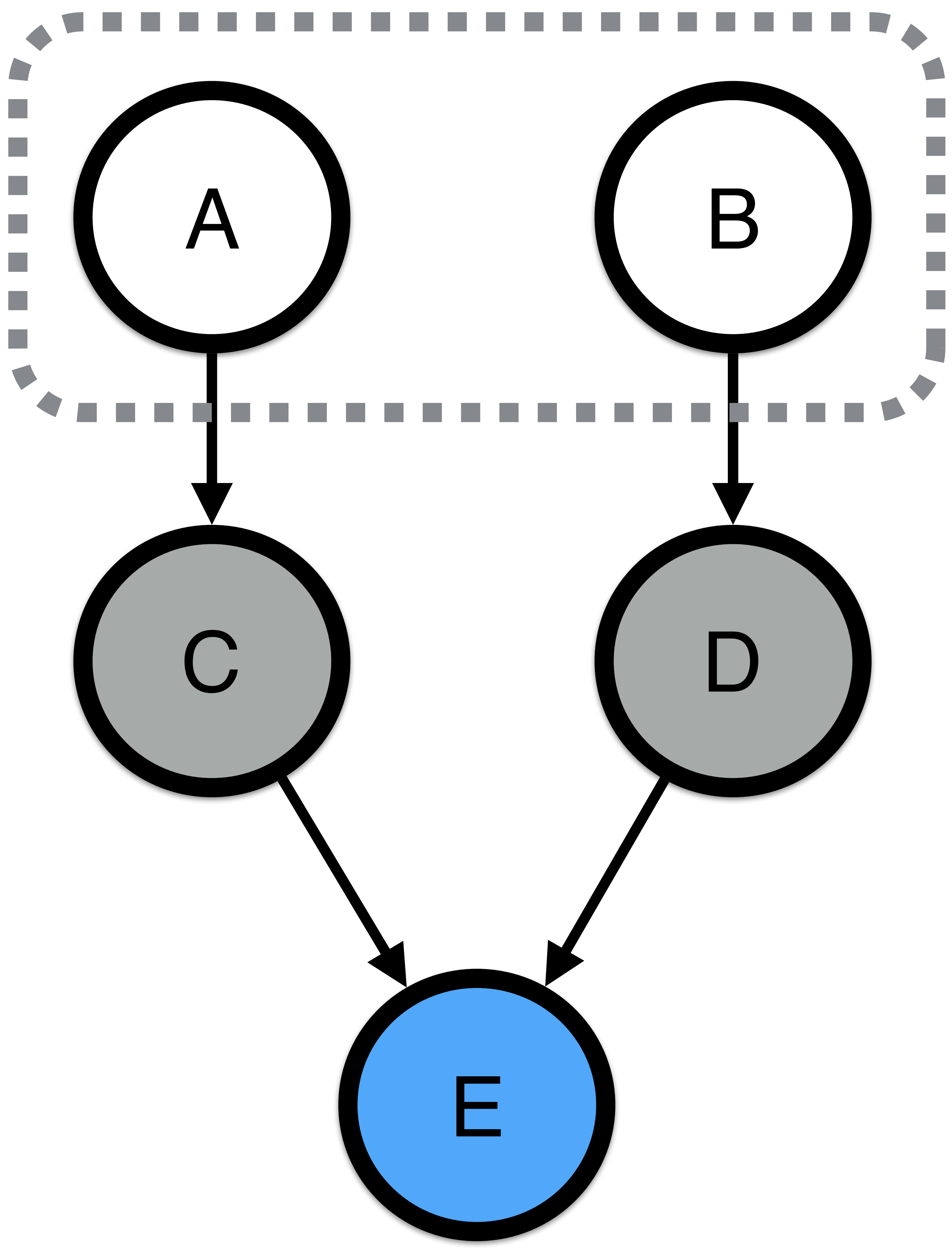
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**



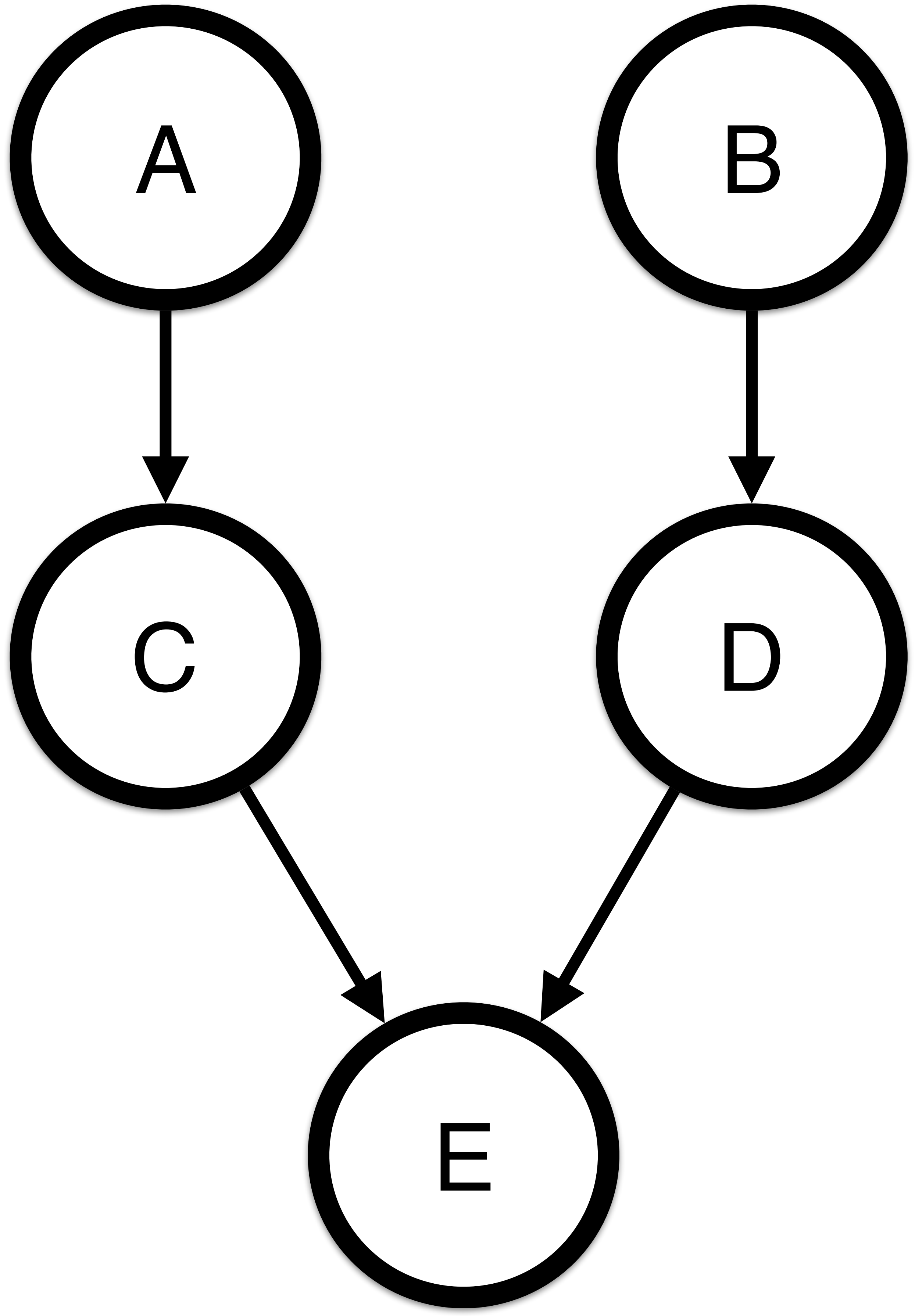
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**



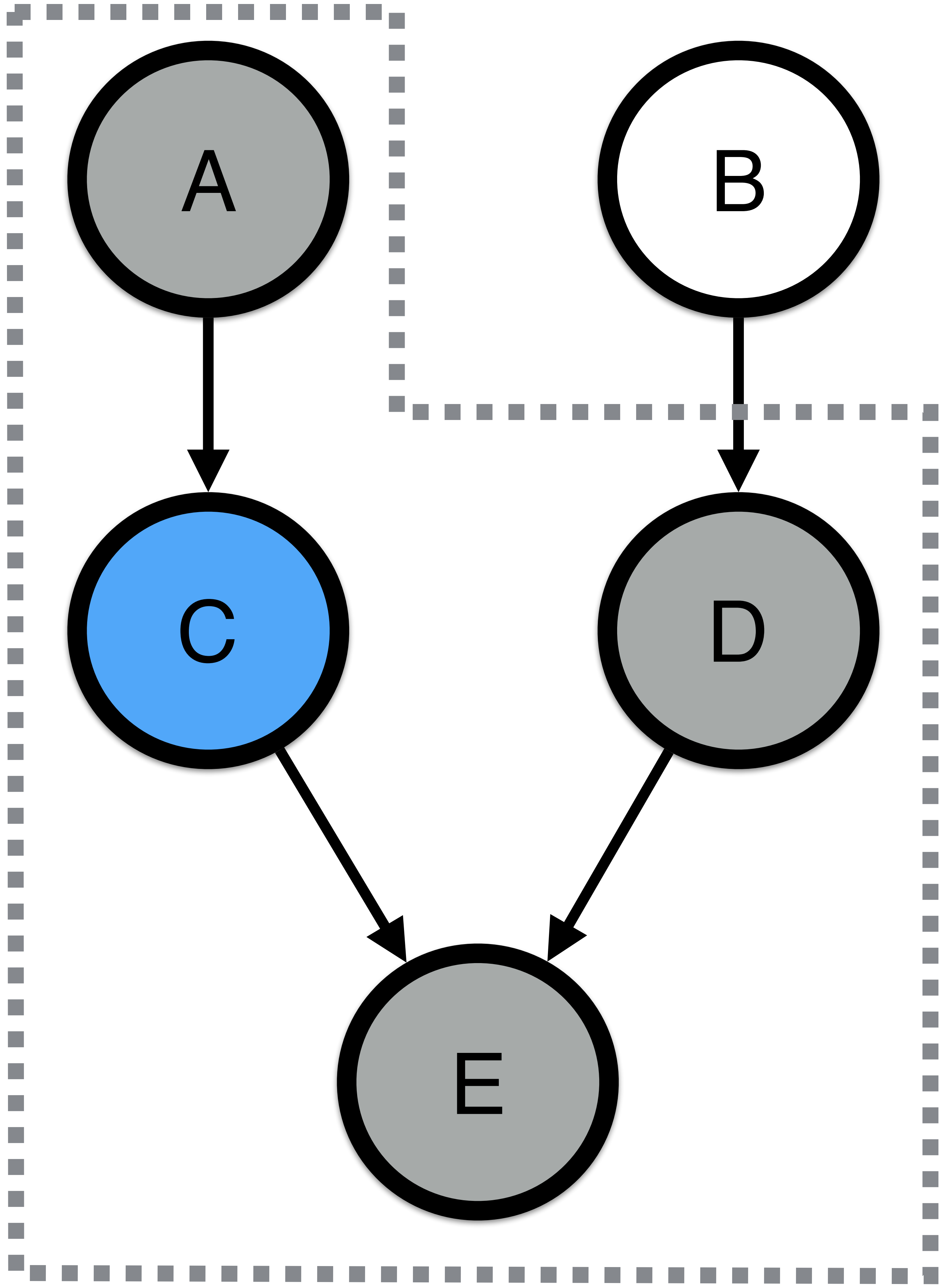
Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**
- Each variable is conditionally independent of any other variable given its **Markov blanket**
  - Parents, children, and children's parents



Independence in Bayes Nets

- Each variable is conditionally independent of its **non-descendents** given its **parents**
- Each variable is conditionally independent of any other variable given its **Markov blanket**
  - Parents, children, and children's parents



Inference

- Given a Bayesian Network describing $P(X, Y, Z)$, what is $P(Y)$
  - First approach: **enumeration**

P(R, W, S, C) = P(R) P(C) P(W | C, R) P(S | W)

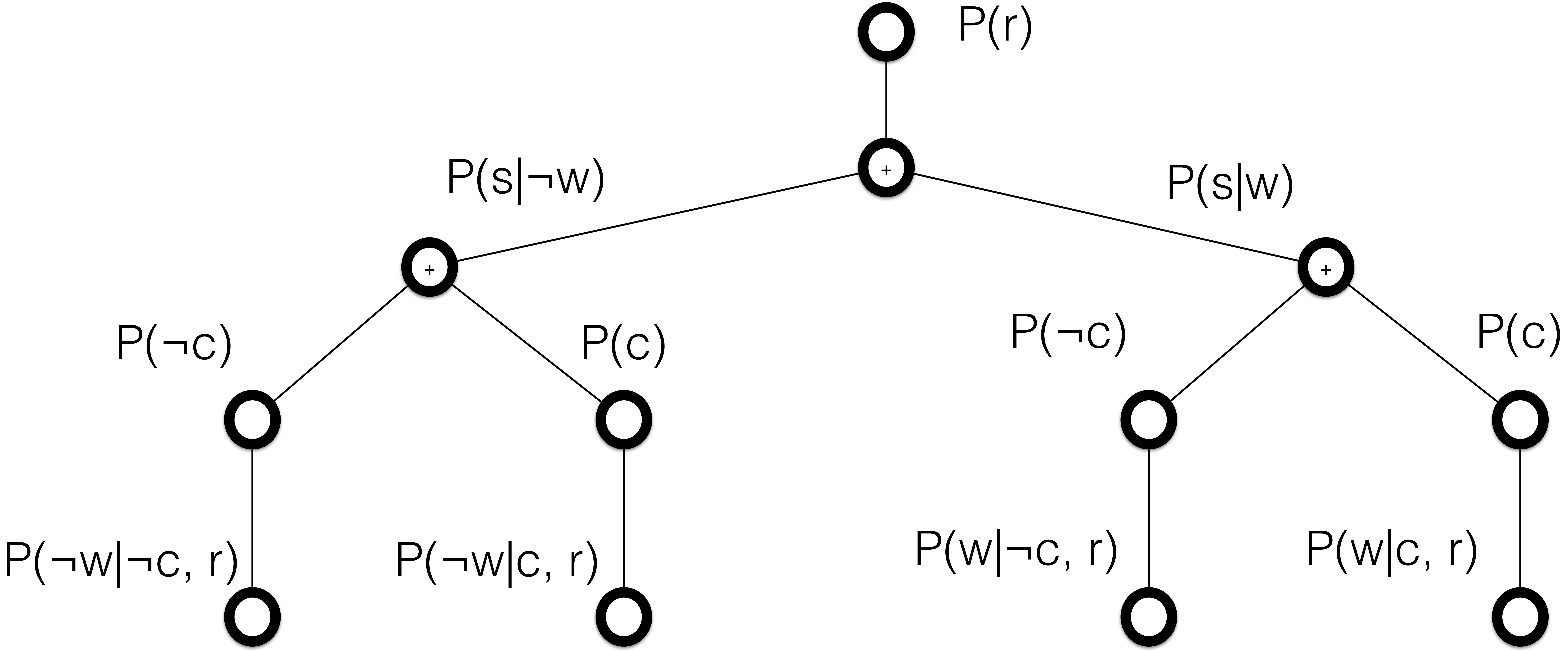
$$P(r|s) = \sum_{w} \sum_{c} P(r, w, s, c)/P(s)$$

$$P(r|s) \propto \sum_{w} \sum_{c} P(r)P(c)P(w|c, r)P(s|w)$$

$$P(r|s) \propto P(r) \sum_{w} P(s|w) \sum_{c} P(c)P(w|c, r)$$

$$O(2^n)$$

$$P(r|s) \propto P(r) \sum_{w} P(s|w) \sum_{c} P(c) P(w|c,r)$$



Second Approach: Variable Elimination

$$P(r|s) \propto \sum_w \sum_c P(r) P(c) P(w|c, r) P(s|w)$$

$$f_C(w) = \sum_c P(c) P(w|c, r)$$

$$P(r|s) \propto \sum_w P(r) P(s|w) f_c(w)$$

$$P(W, X, Y, Z) = P(W)P(X|W)P(Y|X)P(Z|Y)$$

$P(Y)$?

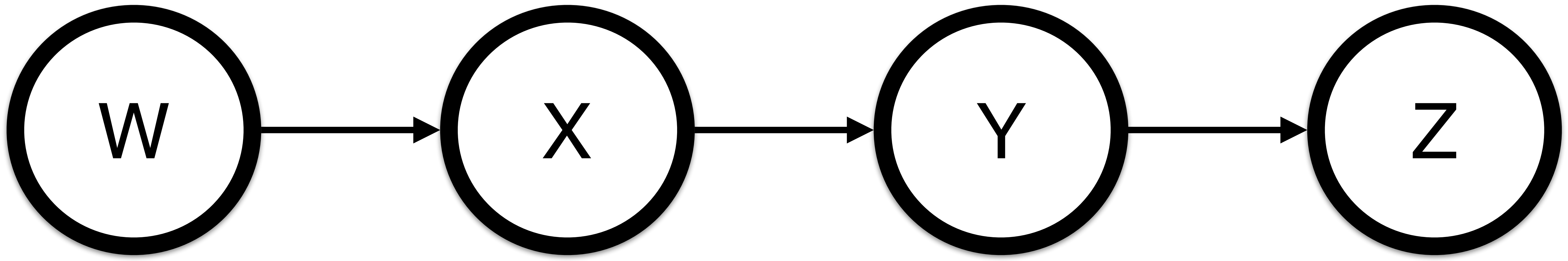
$$P(Y) = \sum_w \sum_x \sum_z P(w)P(x|w)P(Y|x)P(z|Y)$$

$$f_w(x) = \sum_w P(w)P(x|w)$$

$$P(Y) = \sum_x \sum_z f_w(x)P(Y|x)P(z|Y)$$

$$f_x(Y) = \sum_x f_w(x)P(Y|x)$$

$$P(Y) = \sum_z f_x(Y)P(z|Y)$$



$$P(Y) = \sum_w \sum_x \sum_z P(w)P(x|w)P(Y|x)P(z|Y)$$

$$f_w(x) = \sum_w P(w)P(x|w)$$

$$P(Y) = \sum_x \sum_z f_w(x)P(Y|x)P(z|Y)$$

$$f_x(Y) = \sum_x f_w(x)P(Y|x)$$

$$P(Y) = \sum_z f_x(Y)P(z|Y)$$

Variable Elimination

- Every variable that is not an ancestor of a query variable or evidence variable is irrelevant to the query
- Iterate:
  - choose variable to eliminate
  - sum terms relevant to variable, generate new factor
  - until no more variables to eliminate
- Exact inference is #P-Hard
  - in tree-structured BNs, linear time (in number of table entries)

Learning in Bayes Nets

- Super easy!
- Estimate each conditional probability
  - just like we did for naive Bayes