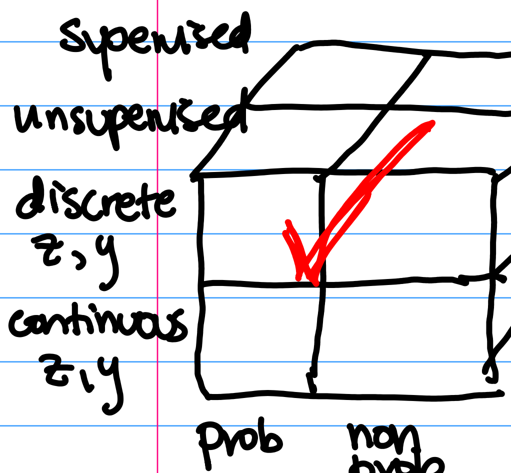
CS181 – Models with Structure (Ch. 8)



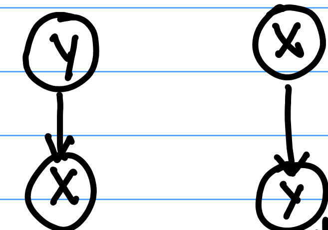
+ model selection, model classes, (NN) objectives (SVM)

+ structured models, decision-making (RL)

Notes: • HW5 due Friday
• Practical teams due Friday (and look at sample code!)

Graphical Models

→ Early: Supervised Learning



generative classifier

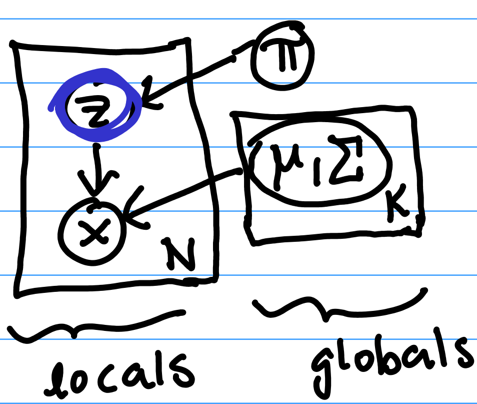
discrim. classifier

$p(x,y) = p(y)\,p(x|y)$

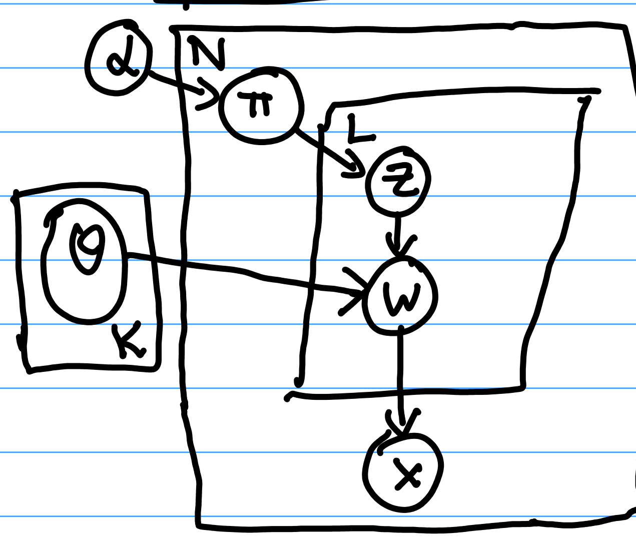
$p(x,y) = p(x)\,p(y|x)$

→ More recently:

Mixture models:



Topic Models:



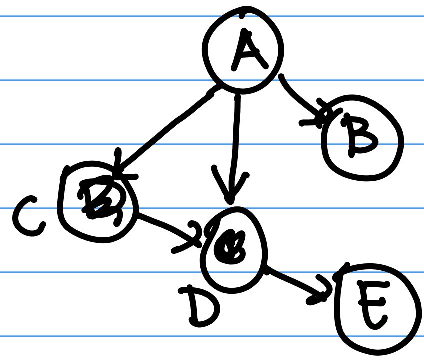
These are examples of __Bayesian__ __Networks__

→ help encode structure of the data (including independence relationships)

→ independences are useful for:

1. inference (block coordinate ascent)
2. learn smaller models

Today: focus __Directed__ __Acyclic__ __Graph__ (DAG)

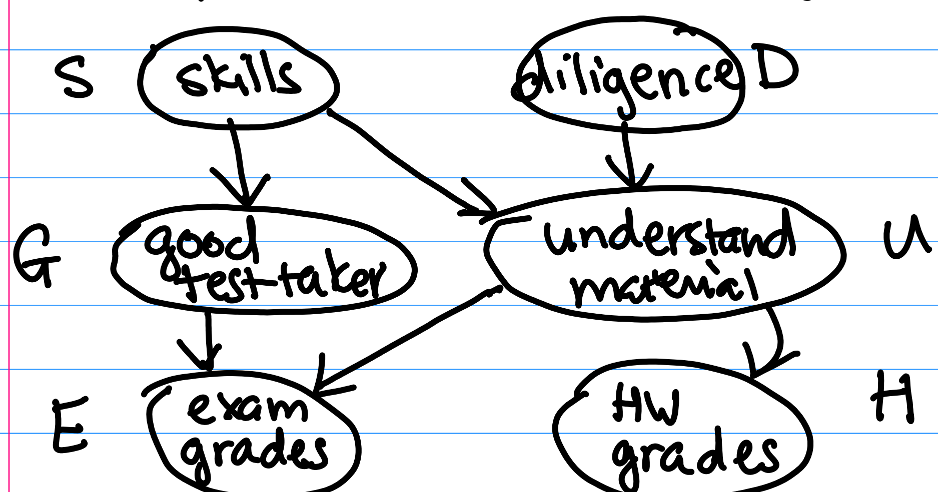


→ interpretation is that we can simplify the joint prob in a specific way:

$$p(A,B,C,D,E) = p(A)\,p(B|A)\,p(C|B,A)\,p(D|A,B,C)\,p(E|A,B,C,D)$$

$$= p(A)\,p(B|A)\,p(C|A)\,p(D|C,A)\,p(E|D)$$

(*) Local independence: every node is conditionally indep of non-descendants given parents



→ given G, U: E is indep of S, D, H

→ given S, are G, U indep?

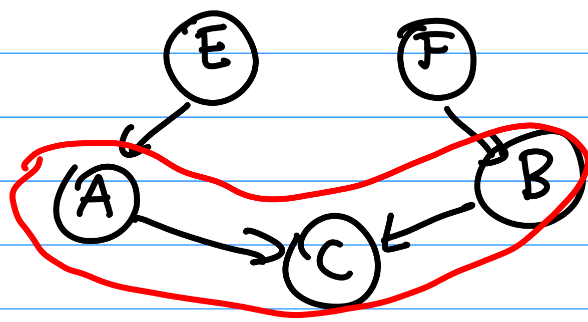
→ given E, are G, U indep?

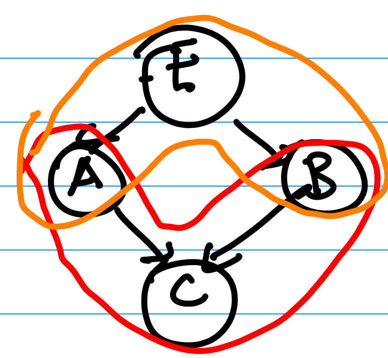
Formalize into rules: "<u>D-separation</u>"

A, B are d-separated if every undirected path from A to B is blocked.

Ways to block:

1. $A \to C \to B$, C is observed
2. $A \leftarrow C \leftarrow B$, C is observed
3. $A \leftarrow C \to B$, C is observed
4. $A \to C \leftarrow B$, C is NOT observed





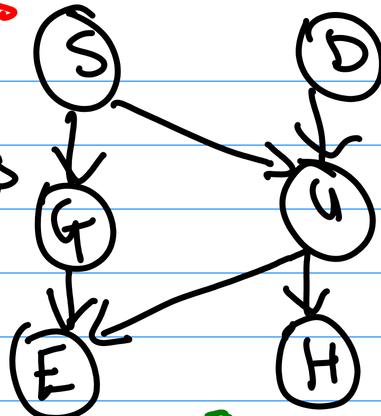
Quick note: about uniqueness:

- if we have a <u>causal interpretation</u> $A \to B$ then A causes B.
- BUT: in general, statistical interpretation allows for multiple orderings / graphs (but one may convey the fewest params / most independences)

$$p(A,B) = p(A)\,p(B|A) = p(B)\,p(A|B)$$

p(s is true)

minimal ordering, most independences



U|S=0,D=0
U|S=0,D=1
U|S=1,D=0
U|S=1,D=1

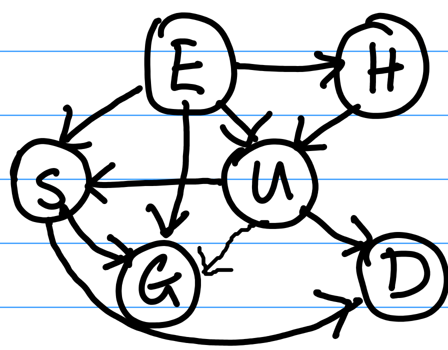
Count the # of params (if we assume all nodes are binary):

S: 1 D: 1 E: 4
G: 2 U: 4 H: 2

Total: 14 params

$p(S)\,p(D|\cancel{S})\,p(G|S,\cancel{D})\,p(U|\cancel{G},S,D)\,p(E|\cancel{S},\cancel{D},G,U)\,p(H|-)$

What if I drew the graph differently?



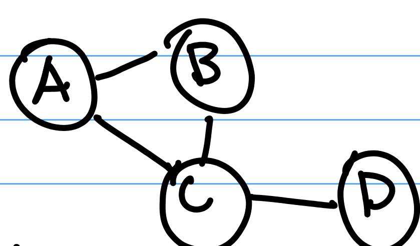
This is a graph with a diff topological ordering, minimal # of connections given that ordering.

count params: 23 params (more than 14)

$p(E)\,p(H|E)\,p(U|E,H)\cdots$

Final notes on other graphical models:

Undirected

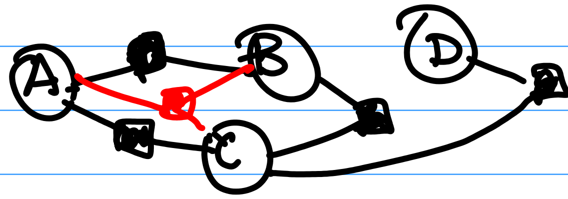


$p(A,B,C,D) =$

$\phi(A,B,C)\,\phi(C,D)$

tricky: normalize??

Factor Graphs



$p(A,B,C,D) =$

$\phi(A,C)\,\phi(A,B)\,\phi(B,C)\,\phi(C,D)$