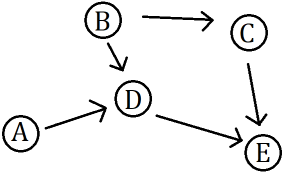
15. Graphical models and HMMs

- graphical models
- HMM

Graphical models

Directed graphs

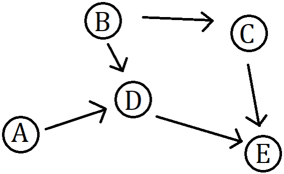


- A directed graph is a collection of
  - nodes (vertices) and
  - directed edges (arcs) between pairs of vertices
- A path is a sequence $V_1, V_2, V_3 ...$ where there exists edges

$$V_1 \to V_2 \to V_3 \to ...$$

- The graph is acyclic if no paths start and end on the same node
- A directed acyclic graph is often just called DAG

Modeling joint distributions via graphical models



- Nodes are random variables
- An arc $X \to Y$ means Y depends on X, and $\Pr(Y|X)$ is available

Example

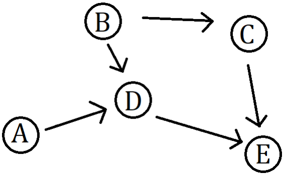
- In figure, E only gets edges from C and D, so

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

  e.g. given C, D, E does not depend on A, B

- Note that this does <u>not</u> mean that $\Pr(E|A,B) = 0$!

Modeling joint distributions via graphical models



Inference

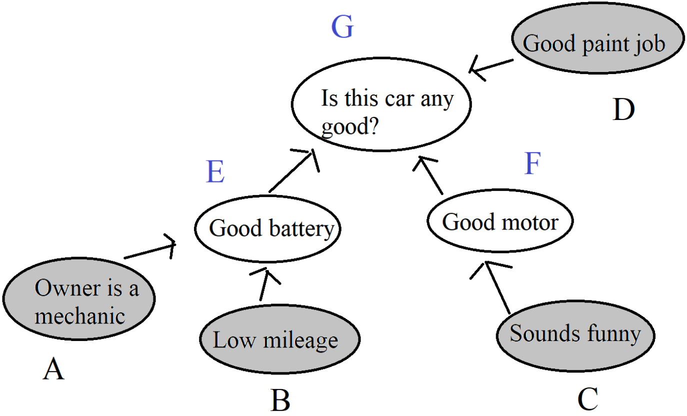
- $A, B, C, D, E \in \mathcal{X}$
- Goal: calculate $\Pr(E) = \int_{\mathcal{X}^4} \Pr(A, B, C, D, E) d(A, B, C, D)$
- Joint probability $\Pr(A, B, C, D, E)$

$$\underbrace{\Pr(E|D, C, B, A)}_{\mathcal{X}^5} \underbrace{\Pr(D|C, B, A)}_{\mathcal{X}^4} \underbrace{\Pr(C|B, \cancel{A})}_{\mathcal{X}^3} \underbrace{\Pr(\cancel{B}|A)}_{\mathcal{X}^2} \underbrace{\Pr(A)}_{\mathcal{X}}$$

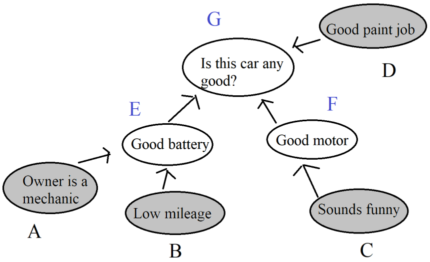
$$= \underbrace{\Pr(E|D, C)}_{\mathcal{X}^3} \underbrace{\Pr(D|B, A)}_{\mathcal{X}^3} \underbrace{\Pr(C|B)}_{\mathcal{X}^2} \underbrace{\Pr(A)}_{\mathcal{X}}$$

- If $\mathcal{X}$ is large, $\mathcal{X}^5 \to \mathcal{X}^3$ is a huge reduction

Example: Should I buy this car?



Example: Should I buy this car?

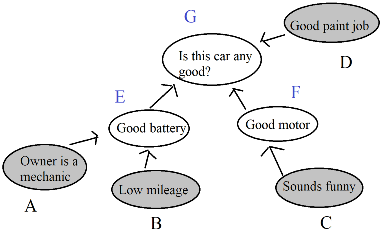


Pr(E | A, B)	A	B
90%	1	1
60%	1	0
60%	0	1
30%	0	0

Pr(F | C)	C
10%	1
70%	0

Pr(G | D, E, F)	D	E	F
90%	1	1	1
10%	1	1	0
60%	1	0	1
5%	1	0	0
89%	0	1	1
9%	0	1	0
58%	0	0	1
4%	0	0	0

Example: Should I buy this car?



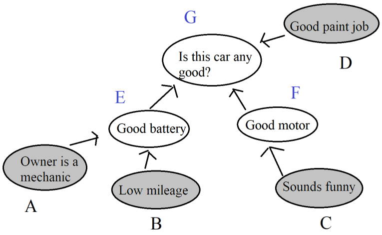
Pr(E | A, B)	A	B
90%	1	1
60%	1	0
60%	0	1
30%	0	0

Pr(F | C)	C
10%	1
70%	0

Pr(G | D, E, F)	D	E	F
90%	1	1	1
10%	1	1	0
60%	1	0	1
5%	1	0	0
89%	0	1	1
9%	0	1	0
58%	0	0	1
4%	0	0	0

- A, B, C, D are observed
- E, F, G are hidden
- I want to infer G

Example: Should I buy this car?



Pr(E | A, B)	A	B
90%	1	1
60%	1	0
60%	0	1
30%	0	0

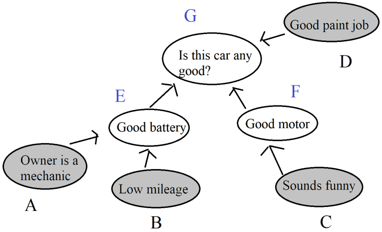
Pr(F | C)	C
10%	1
70%	0

Pr(G | D, E, F)	D	E	F
90%	1	1	1
10%	1	1	0
60%	1	0	1
5%	1	0	0
89%	0	1	1
9%	0	1	0
58%	0	0	1
4%	0	0	0

$$\begin{aligned}\Pr(G|A,B,C,D) &= \sum_{E,F\in\{0,1\}} \Pr(G|A,B,C,D,E,F)\Pr(E,F|A,B,C,D)) \\ &= \sum_{E,F\in\{0,1\}} \Pr(G|D,E,F)\Pr(E|A,B)\Pr(F|C)\end{aligned}$$

Example: Should I buy this car?

$$A = \text{ True}, B = \text{ False}, C = \text{ False}, D = \text{ True},$$



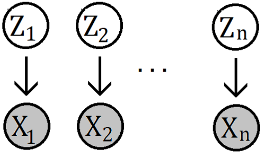
Pr(E | A, B)	A	B
90%	1	1
60%	1	0
60%	0	1
30%	0	0

Pr(F | C)	C
10%	1
70%	0

Pr(G | D, E, F)	D	E	F
90%	1	1	1
10%	1	1	0
60%	1	0	1
5%	1	0	0
89%	0	1	1
9%	0	1	0
58%	0	0	1
4%	0	0	0

$$\begin{aligned}
&\Pr(G|A=1,B=0,C=0,D=1) \\
&= \sum_{E,F\in\{0,1\}} \Pr(G|D=1,E,F)\Pr(E|A=1,B=0)\Pr(F|C=0) \\
&= \underbrace{0.9\cdot 0.6\cdot 0.7}_{E=1,F=1} + \underbrace{0.1\cdot 0.6\cdot 0.3}_{E=1,F=0} + \underbrace{0.6\cdot 0.4\cdot 0.7}_{E=0,F=1} + \underbrace{0.05\cdot 0.4\cdot 0.3}_{E=0,F=0} = 0.498
\end{aligned}$$

Gaussian mixture model as graphical model



- $\Pr(Z_k) = \alpha_k$ mixture weights
- $\Pr(X_k|Z_k) = f_{\mathcal{N}}(X_k; \mu_k, C_k)$ Gaussian distribution
- Inference:

$$\Pr(X_k) = \sum_{Z_k \in \{0,1\}} \Pr(X_k|Z_k)\Pr(Z_k)dZ_k$$

Hidden Markov models

Hidden Markov models (HMMs)

- A random process $A_1, A_2, ...$ has Markov property if

$$\Pr(A_t | A_1, ..., A_{t-1}) = \Pr(A_t | A_{t-1})$$

$$A_1 \to A_2 \to \cdots \to A_T$$

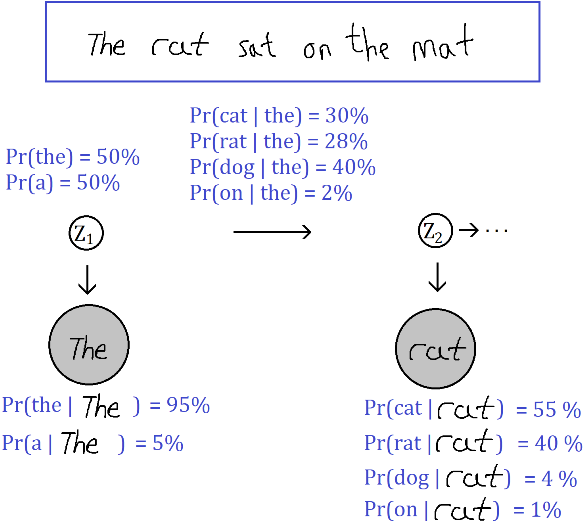
- A hidden Markov model (HMM) divides nodes to two types: hidden (Markovian) and observed (independent given hidden)

$$\begin{tikzcd}
Z_1 \arrow[r] \arrow[d] & Z_2 \arrow[r] \arrow[d] & \cdots \arrow[r] & Z_T \arrow[d] \\
X_1 & X_2 & \cdots & X_T
\end{tikzcd}$$

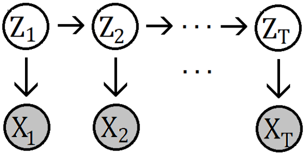
- Before deep learning*, max likelihood estimation over HMMs were state of the art for speech recognition and sequential processing

* e.g. bidirectional LSTMs

HMM example for handwriting recognition

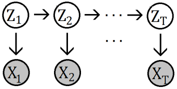


Hidden Markov model distribution



$$
\begin{aligned}
\Pr(X, Z) &\propto \Pr(X|Z)\Pr(Z) \\
&= \underbrace{\Pr(Z_1)\prod_{t=2}^{T}\Pr(Z_t|Z_{t-1})}_{\text{Markov assumption}}\prod_{t=1}^{T}\Pr(X_t|Z_t)
\end{aligned}
$$

Hidden Markov model distribution



$$\Pr(X, Z) \propto \Pr(Z_1) \prod_{t=2}^{T} \Pr(Z_t | Z_{t-1}) \prod_{t=1}^{T} \Pr(X_t | Z_t)$$

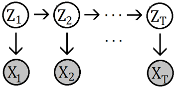
Know:

- Transition probabilities $\Pr(Z_t | Z_{t-1})$ (e.g. prob. that “rat” follows “the”)
- Emission probabilities $\Pr(X_t | Z_t)$ (e.g. prob that  is “the”)
- Initial distribution $\Pr(Z_1)$ (e.g. prob. that “the” is in text)

Want: $\Pr(Z_t | X) =$ probability distribution for tth word

Question: Storage size for $\Pr(X, Z)$ (X given)?

Hidden Markov model distribution



$$\Pr(X, Z) \propto \Pr(Z_1) \prod_{t=2}^{T} \Pr(Z_t | Z_{t-1}) \prod_{t=1}^{T} \Pr(X_t | Z_t)$$

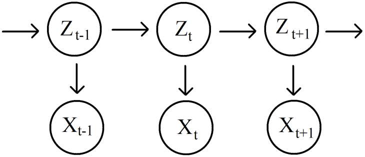
Know:

- Transition probabilities $\Pr(Z_t | Z_{t-1})$ (e.g. prob. that “rat” follows “the”)
- Emission probabilities $\Pr(X_t | Z_t)$ (e.g. prob that  is “the”)
- Initial distribution $\Pr(Z_1)$ (e.g. prob. that “the” is in text)

Want: $\Pr(Z_t | X) =$ probability distribution for tth word

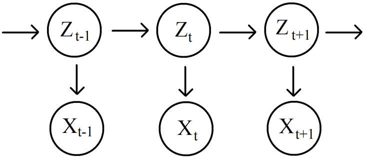
Question: Storage size for $\Pr(X, Z)$ (X given)? Ans: K^T, too many!

Forward-backward (Baum-Welch) algorithm



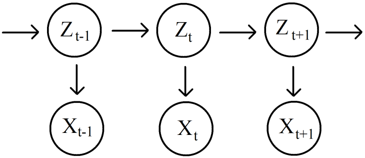
$$
\begin{aligned}
\Pr(Z_t|X) &\propto \Pr(Z_t, X) \\
&= \underbrace{\Pr(X_{t+1:T}|Z_t, X_{1:t})}_{\text{Cond'l indep. assumpt.}} \Pr(Z_t, X_{1:t}) \\
&= \underbrace{\Pr(X_{t+1:T}|Z_t)}_{\substack{\text{Backward} \\ \text{message}}} \underbrace{\Pr(Z_t, X_{1:t})}_{\substack{\text{Forward} \\ \text{message}}}
\end{aligned}
$$

Forward step



$$
\begin{aligned}
\Pr(Z_t, X_{1:t}) &= \sum_{i=1}^{K} \Pr(Z_t, X_{1:t}, Z_{t-1} = k) \\
&= \sum_{i=1}^{K} \underbrace{\Pr(X_t|Z_t)}_{\text{emission}} \underbrace{\Pr(Z_t|Z_{t-1} = k)}_{\text{transition}} \underbrace{\Pr(Z_{t-1} = k, X_{1:t-1})}_{\text{recursion}} \\
\Pr(Z_1, X_1) &= \Pr(Z_1)\Pr(X_1|Z_1)
\end{aligned}
$$

Backward step



$$
\begin{aligned}
\Pr(X_{t+1:T}|Z_t) &= \sum_{i=1}^{K} \Pr(X_{t+1:T}, Z_{t+1} = k|Z_t) \\
&= \sum_{i=1}^{K} \underbrace{\Pr(X_{t+1}|Z_{t+1} = k)}_{\text{emission}} \underbrace{\Pr(Z_{t+1} = k|Z_t)}_{\text{transition}} \underbrace{\Pr(X_{t+2:T}|Z_{t+1} = k)}_{\text{recursion}} \\
\Pr(X_T|Z_T) &= \text{known emission}
\end{aligned}
$$