- ~~g~~ still need to group on CMS.

<fix Piazza; README on wiki>

Mixture models:

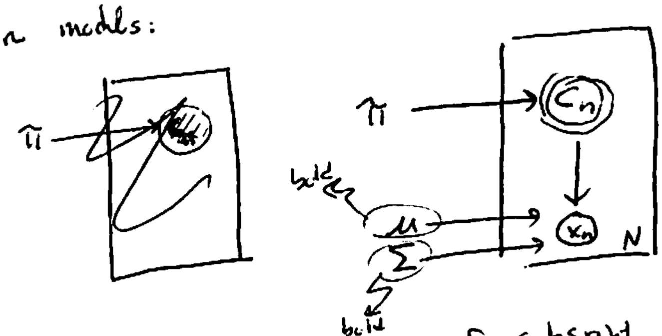


plate: for compactness. n for subscript, N = # draws, or sort of the range
whether something is a circle or not: no circle for parameters w/ no distribution (a parameter). π, μ, Σ are not circled.

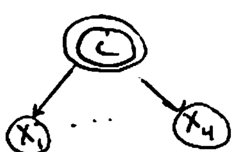
<in our generative story, $\mu = (\mu_1, \ldots, \mu_K)$>
shaded on slides = double circle in my handwritten, shading means "hidden".
~~(note: mismatch~~

Expanding the plate diagram to get the distribution:
~~each~~ the parent of each x_n is c_n.: so, $P(x_n|c_n) \sim N(\vec{\mu}_k, \boxed{\Sigma}_k)$

we also know $p(c_n = l) = \pi[l]$

q: why no edge btwn c & μ?
a: you have to keep straight that the arrows in graph models is not data flow; it's about what ~~g~~ affects which var's distribution.

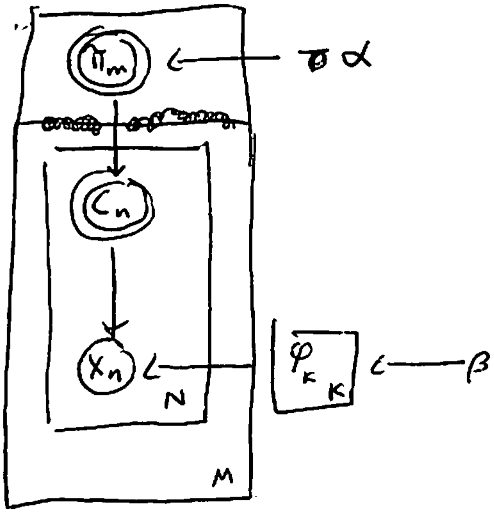
example: Naive Bayes.



$x_i \perp x_j \mid c$.

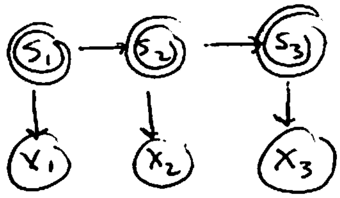
e.g., p("lottery"|spam) is high, p("lottery"|not spam) is low.

LDA:



(talked about π as documents)

HMM:



Kalman filters = both types continuous
(e.g. tracking baseball players)
X_i: frames.
~~stat~~ S_j: location of player.
~~Your location at~~
given your location @ time t,
~~doesn't~~ location @ time $t-1$
doesn't matter.

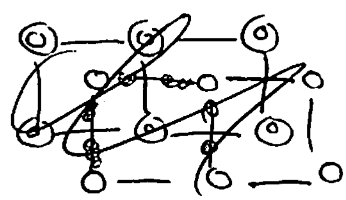
- - - - - - - - - - - -

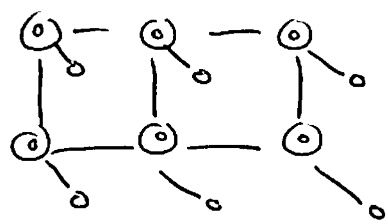
- things that can't be represented as prob distributions:
  Markov random fields / undirected graphical models.
  ~~Can be~~ <and vice versa: →

ex:

no undir can represent this. (you would have to draw a △ for the undir case, b/c X_2 & X_3 can explain X_1)

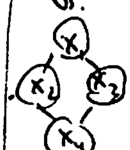
ex: images:





the ⊙ are like, what object is "generating" the pixel value

vs.



the cyclic nature can't be in a ~~gra~~ dir. graphical model.

in MRFs: • non-adj. variables are cond. indep. given all the other vars
• given its nbrs, a variable is cond. indep of all its other variables
• any 2 sets of variables are cond indep given a separating set.

powerful: ~~gre~~ give nice methods
many of our methods will work in this more general setting.

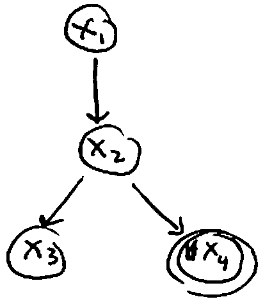
and there's stuff neither can capture

how about inference – assuming parameters fixed, can we answer:

~~- what's~~

marginal: what's the prob of a given assignment for a subset of variables?

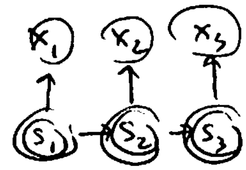
~~cont~~



ex: what is $P_\theta(x_4=1)$?

conditional: given a particular assignment (observation) for some vars, can we predict some other variables?

ex: in HMM:



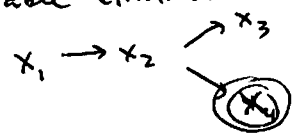
given x_1, what is $P_\theta(s_2 \mid x_1 =)$?

b/c of Bayes' rule, lucky for us, we only need to compute marginals.

~~ex~~ $P(s_3 \mid x_1, x_2, x_3) = ?$

$$= \frac{P(s_3, x_1, x_2, x_3)}{P(x_1, x_2, x_3)} \quad \begin{matrix} \leftarrow \text{marginal} \\ \leftarrow \text{marginal} \end{matrix}$$

ex: variable elimination: given a Bayes Net.



assume params known, and we want to know $P(x_4 \mid x_1, x_2, x_3)$

$$\sum_{x_1, x_2, x_3} P(x_4, x_1, x_2, x_3)$$ <marginalization>

let's factorize using the graph (each var given its parents)

$$= \sum_{x_1 x_2 x_3} \underbrace{P(x_4 \mid x_2) P(x_3 \mid x_2) P(x_2 \mid x_1) P(x_1)}$$

Now, how to estimate the sum?
For simplicity, let's assume binary variables:

$$= \underbrace{\sum_{x_3 \in \{0,1\}} \sum_{x_2 \in \{0,1\}} \sum_{x_1 \in \{0,1\}}} \Big[\qquad \Big]$$

Pick ~~ord~~ an arbitrary ordering on the remaining vars.

$$= \sum_{x_3 \in \{0,1\}} \sum_{x_2 \in \{0,1\}} P(x_3|x_2)P(x_4|x_2) \underbrace{\sum_{x_1 \in \{0,1\}} P(x_1)P(x_2|x_1)}$$

let's call this $m_{x_1}(x_2)$, to note the var I eliminated.

this is the marginal prob of x_2, hence "m".

Note that this is a sum of products.

$$= \sum_{x_3 \in \{0,1\}} m_{x_2}(x_3, x_4)$$

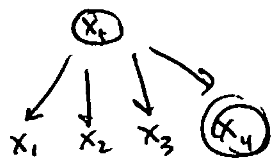
since the thing we collapse is a fn of x_3 & x_4.

and this is a dynamic program.

Preview of belief propagation.

.. it turns out above that picking a good order is critical.

ex.



2 possible orderings: x_5 innermost,

x_5 outermost $\leftarrow$ best.

(how could you eliminate x_5?)