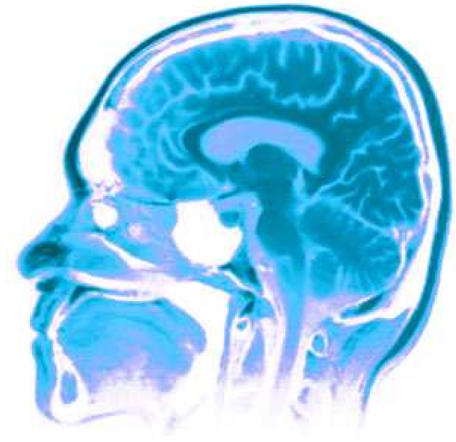
CPSC340



Learning Bayesian Nets

Nando de Freitas
October, 2012
University of British Columbia

Outline of the lecture

This lecture is about applying **Frequentist learning and Bayesian learning** to learn the parameters of directed probabilistic graphical models. The goal is for you to learn:

- ❑ How to apply **maximum likelihood** so as to learn the parameters of the conditional probability tables from data.

- ❑ How to apply **Bayesian learning** with Beta priors and Bernoulli likelihoods to compute the posterior distribution of all the parameters of a Bayesian network.

Learning Bayes nets

*Suppose we are given a **dataset** indicating whether you drank a martini (**M**), whether you went to Fritz for fries after (**F**), whether you stayed home studying (**S**) and whether you got thin (**T**) as a result.*

M	F	S	T
1	1	0	0
1	0	1	1
1	1	0	1
1	1	1	1
0	0	0	0

$n = 5$ observations

Learning Bayes nets

Next, we choose a model describing how we believe each variable influences the other.

$\theta = P(M=1)$

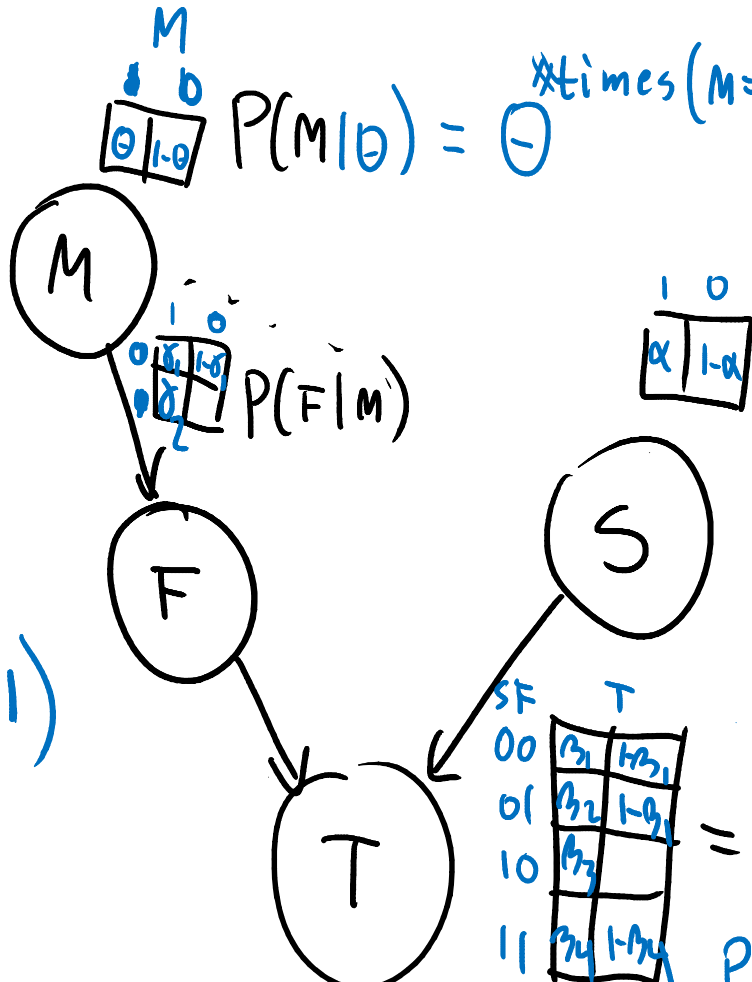
$\beta_1 = P(T=1 \mid F=0, S=0)$

$\gamma_1 = P(F=1 \mid M=0)$

$\gamma_2 = P(F=1 \mid M=1)$

$P(M \mid \theta) = \theta^{\#\text{times}(M=1)} (1-\theta)^{\#\text{times}(M=0)}$

$P(S) = \alpha^{\#t(S=1)} (1-\alpha)^{\#t(S=0)}$



$P(T \mid S=0, F=0) = \beta_1^{\#t(T=1, S=F=0)} (1-\beta_1)^{\#t(T=0, F=0, S=0)}$

Learning Bayes nets

Given the binary observations, we use ***Bernoulli*** *distributions to describe the probabilities of each of the variables in the Bayes net.*

$$P(M \mid \theta) = \theta^4 (1-\theta)^1$$

$$P(TMSF) = P(T \mid F, S) P(F \mid M) P(M) P(S)$$

$$P(S_{1:5} \mid \alpha) = \alpha^2 (1-\alpha)^3$$

$$P(F \mid M=0, \gamma_1) = \gamma_1^0 (1-\gamma_1)^1$$

$$P(F \mid M=1, \gamma_2) = \gamma_2^3 (1-\gamma_2)^1$$

$$P(T \mid F=0, S=0, \beta_1) = \beta_1^0 (1-\beta_1)^1$$

$$\vdots$$

Maximum likelihood for Bayes nets

The maximum likelihood estimates are simply the frequency counts.

$$\hat{\theta}_{ML} = \frac{\#\ 1\text{'s}}{\#\ \text{tries}} = \frac{4}{5}$$

$$\hat{\alpha}_{ML} = \frac{2}{5}$$

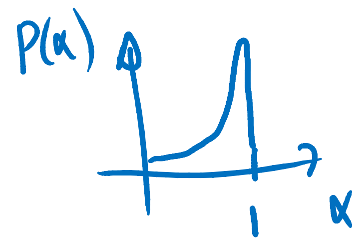
$$\hat{\gamma}_{1_{ML}} = 0/1 = 0$$

$$\hat{\beta}_{1_{ML}} = 0/1 = 0$$

Bayesian learning for Bayes nets

We specify ***Beta priors*** *for each of the variables. Then, we multiply these priors times the* ***Bernoulli likelihoods*** *to derive the* ***Beta posteriors.***

$$P(\alpha) = \mathrm{Beta}(10,1) \propto \alpha^{10-1}(1-\alpha)^{1-1} = \alpha^{9}(1-\alpha)^{0}$$



$$P(\alpha \mid S_{1:5}) = P(S_{1:5} \mid \alpha)\, P(\alpha)$$

$$\propto \alpha^{2}(1-\alpha)^{3}\left[\alpha^{9}(1-\alpha)^{0}\right]$$

$$= \alpha^{11}(1-\alpha)^{3} \qquad \bar{\alpha} = E(\alpha \mid S) =$$

$$= \frac{12}{12+4} = \frac{12}{16} = \frac{3}{4} = 0.75$$

Bayesian learning for Bayes nets

*We specify **Beta priors** for each of the variables. Then, we multiply these priors times the **Bernoulli likelihoods** to derive the **Beta posteriors.***

Inference with the learned net

Given the parameters we have learned, what is the probability that you were drinking martinis given that you're not thin? i.e.

$P(M=1|T=0)$ = ?

Exercise

5 extra marks

by MLE. (5 marks)

by Posterior mean (5 marks)

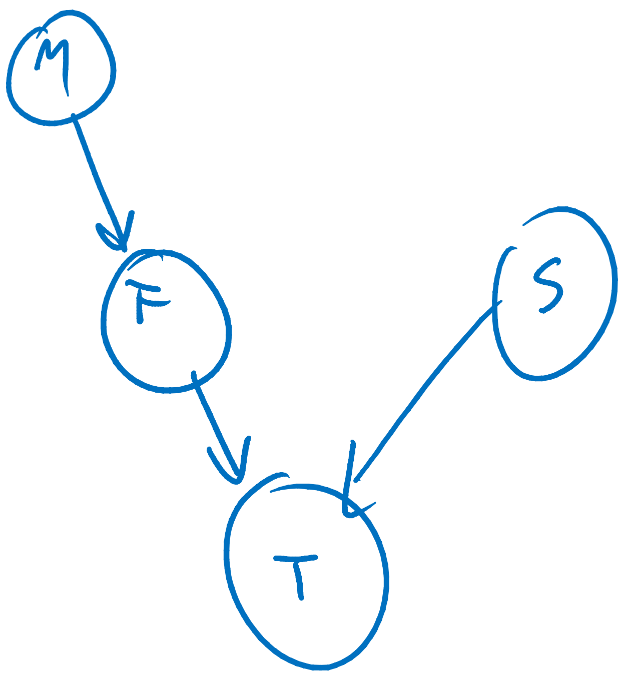
(all priors Beta(1,1))

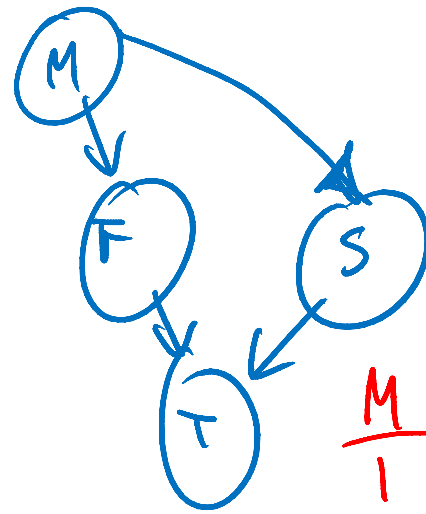
Frequentist model selection

Given a new dataset ***{M,F,T,S}****, we can evaluate the probability of each model structure (using the parameters we learned by maximum likelihood) and pick the model with the highest* ***P(M,F,T,S****|parameters****)****.*

$P(T|FS)P(F|M)P(M)P(S)$

$P(T|FS)P(F|M)P(M)P(S|M)$





M	F	S	T
1	1	0	0
1	0	0	0

For Bayesian model selection, please see the tutorial of David Heckerman on the course website.

Next lecture

In the next lecture, we revise linear algebra and sketch a convergence proof for Google's page rank algorithm.