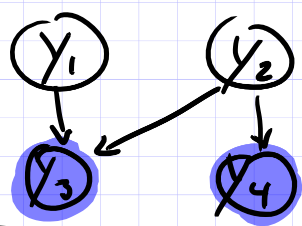
Lecture 23: 4/12/17

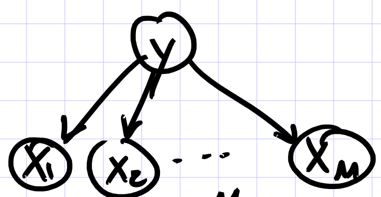
Familiar Models as Bayes Nets

In a graphical model, shaded nodes are "observed".
Their values are given



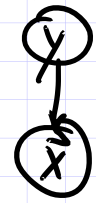
e.g. $P(Y_1, Y_2 \mid Y_3 = 0, Y_4 = 1)$

① Bernoulli Naive Bayes
Gaussian Naive Bayes



$$p(\vec{x}, y) = p(y) \prod_{m=1}^{M} p(x_m \mid y)$$

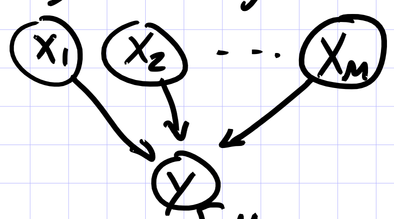
② Gaussian Discriminant Analysis
Gaussian Mixture Model



$$p(\vec{x}, y) = p(y)\, p(\vec{x} \mid y)$$

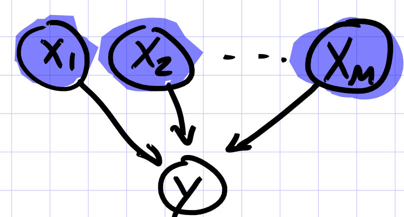
③ Logistic Regression ??

Not



$$p(\vec{x}, y) = \left[\prod_{m=1}^{M} p(x_m)\right] \underbrace{p(y \mid x_1, \ldots, x_M)}_{\text{this part looks like Log. Reg.}}$$

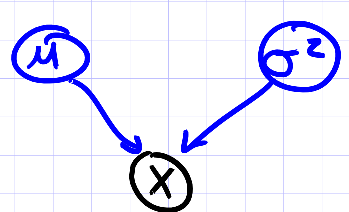
④ Logistic Regression
Linear Regression



$$p(y \mid \vec{x})$$

⑤ 1D Gaussian

$\mu \sim \text{Laplace}(0,1)$
$\sigma^2 \sim \text{Exp}(\cdot\cdot)$
$x \sim \text{Gaussian}(\mu, \sigma^2)$

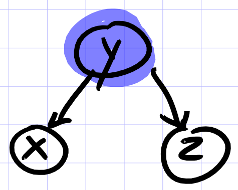


★ parameters can be treated as random variables

there is a lot of info missing from graphical model diagrams. So why use them?

Proofs of Cond. Indep.

Case #1 : Common Parent



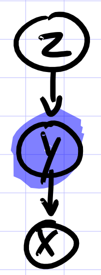
Claim : $X \perp\!\!\!\perp Z \mid Y$

Proof : $$\frac{P(X,Y,Z)}{P(Y)} = \frac{\cancel{P(Y)}\, P(X \mid Y)\, P(Z \mid Y)}{\cancel{P(Y)}}$$

$$\Rightarrow P(X,Z \mid Y) = P(X \mid Y)\, P(Z \mid Y)$$

$$\Rightarrow X \perp\!\!\!\perp Z \mid Y$$

Case #2 : Cascade

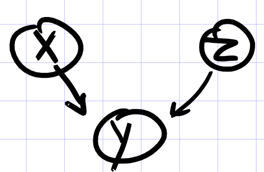


Claim : $X \perp\!\!\!\perp Z \mid Y$

Proof : $$\begin{aligned} P(X,Y,Z) &= \underline{P(Z)\, P(Y \mid Z)}\, P(X \mid Y) \\ &= \underline{P(Z,Y)}\, P(X \mid Y) \\ &= \underline{P(Y)\, P(Z \mid Y)}\, P(X \mid Y) \end{aligned}$$ ← same as case #1

$$\Rightarrow X \perp\!\!\!\perp Z \mid Y$$

Case #3 V-structure



Claim : $X \perp\!\!\!\perp Z$

Proof : $$\begin{aligned} P(X,Y,Z) &= P(X)\, P(Z)\, P(Y \mid X,Z) \\ &= P(X)\, P(Z)\, \frac{P(Y,X,Z)}{P(X,Z)} \end{aligned}$$

$$\Rightarrow P(X,Z) = P(X)\, P(Z)$$

$$\Rightarrow X \perp\!\!\!\perp Z$$

Claim : $X \not\perp\!\!\!\perp Z \mid Y$

Proof : ★ left as an exercise

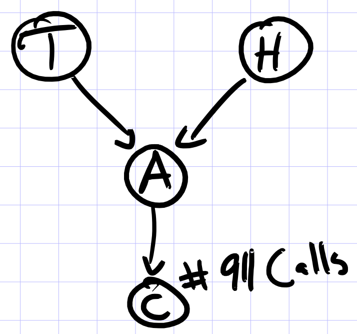
Background : Cond. Indep.

r.v.s X and Z are <u>cond. indep.</u> given a set $Y_1, \ldots, Y_J$ iff

$$P(X,Z \mid Y_1, \ldots, Y_J) = P(X \mid Y_1, \ldots, Y_J)\, P(Z \mid Y_1, \ldots, Y_J)$$

Supervised Learning for Bayes Nets

Ex: Tornado



$T \sim \text{Bernoulli}(\tau)$

$H \sim \text{Bernoulli}(\eta)$

$A \sim \text{Bernoulli}(\alpha_{H,T})$

$C \sim \text{Unif}(\{1,\dots,63\}) + A * \text{Unif}(\{1,\dots,63\})$

integer

MLEs in Closed Form

$$\ell(\tau,\eta,\alpha) = \log \prod_{i=1}^{12} p(t^{(i)}, h^{(i)}, a^{(i)}, c^{(i)} \mid \eta, \tau, \alpha)$$

$$= \sum_{i=1}^{12} \log p(t^{(i)} \mid \tau) + \log p(h^{(i)} \mid \eta) + \log p(a^{(i)} \mid t^{(i)}, h^{(i)}, \alpha) + \log p(c^{(i)} \mid a^{(i)})$$

$$\hat{\tau}, \hat{\eta}, \hat{\alpha} = \operatorname{argmax} \ell(\tau,\eta,\alpha)$$

$$\hat{\tau} = \operatorname*{argmax}_{\tau} \sum_{i=1}^{12} \log p(t^{(i)} \mid \tau) = \#(T=1)/N$$

$$= \#(H=1)/N$$

$$\hat{\eta} = \sim\sim$$

$$\hat{\alpha} = \operatorname*{argmax}_{\alpha} \sum_{i=1}^{12} \log p(a^{(i)} \mid t^{(i)}, h^{(i)}, \alpha)$$

$$\alpha_{h,t} = \frac{\#(A=1, T=t, H=h)}{\#(T=t, H=h)}$$