CSC321 T4

ex 1

Conditional probility table (CPT)

X = M / F . y = smoker / non-smoker

| | | Y: S | NS |
|---|---|---|---|
| X | M | $P_1$ | $P_2$ |
| | F | $P_3$ | $P_4$ |

e.g. $P_1 = P(X=M, Y=\text{smoker})$

$P_1 + P_2 + P_3 + P_4 = 1$

a.

$$P(Y=S) = P(Y=S, X=M) + P(Y=S, X=F)$$

b.

$$P(Y=S \mid X=M)$$

$$= \frac{P(Y=S, X=M)}{P(X=M)}$$

$$= \frac{P(Y=S, X=M)}{P(X=M, Y=S) + P(X=M, Y=NS)}$$

$$= \frac{P_1}{P_1 + P_2}$$

1

ex 2

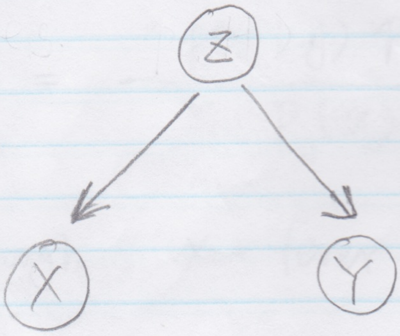
## Conditional independence

Toss coin $X = H$ or $T$ coin $Y = H$ or $T$

$X \perp\!\!\!\perp Y$. $\quad P(X=H, Y=H) = P(X=H)P(Y=H)$

$\qquad\qquad\quad P(Y=H \mid X=H) = P(Y=H)$

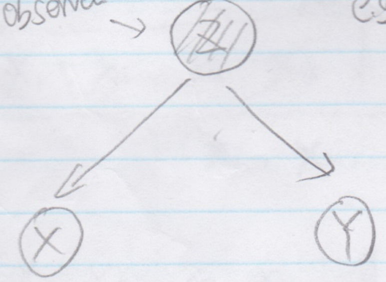
Given $Z$: biased toward $H$ or $T$



$X \not\perp\!\!\!\perp Y$

$$P(Y=H \mid X=H) > P(Y=H)$$

observed



e.g., 0.6 H, 0.4 T

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

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

ex3.

Bayes rule : $p(y \mid x)\, p(x) = p(x, y) = p(x \mid y)\, p(y)$

example: $X$ = cough

$Y$ = lung cancer

$p(Y \mid X) = ?$ not known

Assume we know:

$p(X \mid y)$ , $p(X)$ , $p(y)$

Diagnosis:

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

ex b.

$x_1$ = cough . $x_2$ = fever . $x_3$ = smoke . $x_4$ = air pollution

$y$ = lung cener

$$p(y \mid x_1, x_2, x_3, x_4)$$

$$= \frac{\overbrace{p(x_1, x_2, x_3, x_4 \mid y)}^{2^4 = 16 \text{ possibilities}}\, p(y)}{p(x_1, x_2, x_3, x_4)}$$

$$\propto p(x_1, x_2, x_3, x_4 \mid y)\, p(y)$$

$O(2^D)$ for $D$ variables

$$\begin{array}{c|cccc|c} & \multicolumn{4}{c}{p(x_1, x_2, x_3, x_4 \mid y)} \\ \hline & 1 & 1 & 1 & 1 \\ & 0 & 1 & 1 & 1 \\ 16 & 1 & 0 & 1 & 1 \\ & 1 & 1 & 0 & 1 \\ & & & \vdots & \\ & 0 & 0 & 0 & 0 \end{array}$$

Naïve Bayes Assumption: $x_1 \perp\!\!\!\perp x_2 \mid y$

$$\cong \underbrace{p(x_1 \mid y)\, p(x_2 \mid y)\, p(x_3 \mid y)\, p(x_4 \mid y)}_{4}\, p(y)$$

$$= p(y) \prod_k p(x_k \mid y)$$

$O(D)$ for $D$ variables

ex4. Maximum likelihood (ML) of Gaussian mean

$$p(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$p(x_1, x_2, \dots x_N \mid \mu, \sigma^2) \overset{\text{i.i.d.}}{=\!=} \prod_{i=1}^{N} p(x_i \mid \mu, \sigma^2)$$

i.i.d.: ① independence ② identically distributed

$$L = \log p(\vec{x} \mid \mu, \sigma^2) = \log \prod_{i=1}^{N} p(x_i \mid \mu, \sigma^2)$$

$$= \sum_{i=1}^{N} \log p(x_i \mid \mu, \sigma^2)$$

$$= \sum_{i=1}^{N} \left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) + \text{const}$$

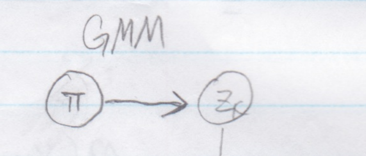
$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{N} (x_i - \mu)$$

$$= \frac{1}{\sigma^2}\left(\sum_{i=1}^{N} x_i - N\mu\right)$$

$\frac{\partial L}{\partial \mu} = 0$ and solve for $\mu$

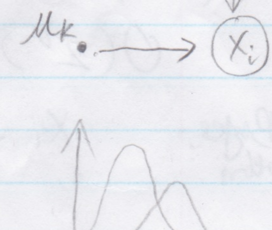
$$\frac{1}{\sigma^2}\left(\sum_{i=1}^{N} x_i - N\mu\right) = 0$$

GMM




$$\mu_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

i.i.d.




Violation of ② : Gaussian mixture model.
e.g. two classes

-- ① : hard problem. possible solution: hidden Markov model
e.g. ① switching b/w biased & unbiased coin: HHHTHTHHTTHTHH....
② virus outbreak