$\boxed{\text{Ex}}$

$P \perp C$? yes.

$P \perp C \mid MIT$?

P   C $P(P \mid C \cdot MIT) < P(P \mid MIT)$

↘ ↙

MIT. $< P(P \mid \bar{C} \cdot MIT)$

$P \not\perp C \mid MIT$

<f.> Explain Toward   prob high.

$P(A \mid C) < P(A \mid BC)$.

Model Selection.

have data / evidence & multiple models

M_i's

→ pick most model (max $P(M_i \mid data)$).

$\boxed{\text{Ex}}$ 2 coins: fair ($P(H) = 0.5$) C_1

unfair ($P(H) = 1$) · C_2

2 models ① [draw coin randomly & flip] × 3

② draw coin once " flip × 3.

data = HHT.

– assume.. prior ⊜ $P(M_1) = P(M_2) = \frac{1}{2}$

MAX $P(M = M_i \mid data)$. $= \dfrac{P(data \mid M = M_i) P(M = M_i)}{P(data)}$ ← const

$P(data)$ ← const.

$M_1 = P(HHT \mid M_1) = \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{9}{64}$.

$P(H) = P(H \mid C_1) \cdot P(C_1) + P(H \mid C_2) \cdot P(C_2) = \frac{1}{2} \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{3}{4}$

$M_2 = P(HHT \mid M_2) =$

$= P(HHT \mid C_1) \cdot P(C_1) + P(HHT \mid C_2) P(C_2) = (\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}) \frac{1}{2} + (1 \cdot 1 \cdot 0) \frac{1}{2}$

$= \frac{1}{16}$.

Application

Naive Bayes Classification

– assumption " features are cond indep. of each other given classification

"given parents"

$Y \to X_1, X_2, X_3$

$\boxed{\text{Ex}}$ "likes toy" inverse "likes veggies".

(for entire pop).

"given child" "toy" ⊥ "veggies".

training pts (data) → calculate prob.

classification: max. $\boxed{P(Y = y \mid data)}$.

posterior prob.

$= \dfrac{P(data \mid Y = y) \boxed{P(Y = y)}}{P(data)}$

prior

const

assumptions

$\propto P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n \mid Y = y)(P(Y = y))$

$\propto P(X_1 = x_1 \mid Y = y) \cdot P(\cdot) \cdots P(X_n = x_n \mid Y = y)$.

$C_1 \to F_1$, $C_2 \to F_2$, $C_3 \to F_3$.

$C_1 \to F_1, F_2, F_3$.