$w \quad f : \langle X \rangle \to Y$

$g \approx f$

└→ Concept Learning

└→ Decision Tree Learning

$\boxed{TE}$ ✓

- $\text{Prob}(Y=1 \mid \langle x \rangle)$?

- $\text{Prob}(Y=0 \mid \langle x \rangle)$? $(1 - ?)$

$\langle x^{new} \rangle \longrightarrow y^{+1, -1}$

✓ Prob (max)

$y=0$

$y=1$

[Classify]

- Random Variable:
  - └→ outcome of Random Exp.

$$\text{Prob}(A=f) = \frac{|S_f|}{|S|} \quad \begin{matrix} \to \text{\# female} \\ \to \text{\# student} \end{matrix}$$

[A → female student]

- **Axioms:** (1) $P(A) + P(\bar{A}) = 1$

(2) $0 \leq P(A) \leq 1$

(3) $P(A \underset{\cup}{\vee} B) = P(A) + P(B) - P(A \underset{\cap}{\wedge} B)$

✓ (4) $P(A) = P(A \wedge B) + P(A \wedge \bar{B})$

$P(A|B).P(B)$ $\qquad$ $P(A|\bar{B})\,P(\bar{B})$

$A = (A \wedge B) \vee (A \wedge \bar{B})$

▷ Conditional Prob: $P(A|B) = \dfrac{P(A \wedge B)}{P(B)}$ ✓

Bayes

$$\Rightarrow P(A \wedge B) = \underline{P(A|B)\, P(B)}$$
$$= \underline{P(B|A)\, P(A)}$$

Premise of Bayes Rule

$$\boxed{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$$

$P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$

$$P(A \wedge B \wedge C) = P(A|B \wedge C)\, P(B \wedge C)$$
$$= P(A|BC)\, P(B|C)\, P(C)$$

$$P(A_1 \wedge A_2 \cdots \wedge A_n) = P(A_1|A_2 \cdots A_n)\, P(A_2|A_3 \cdots A_n)\, P(A_3|A_4 \cdots A_n) \cdots P(A_{n-1}|A_n)\, P(A_n)$$

(Chain Rule)

Generalize Bayes Th-i

(1) $$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|\bar{A})\,P(\bar{A})}$$

(2) $$P(A|B \wedge X) = \frac{P(B|A \wedge X)\,P(A \wedge X)}{P(B \wedge X)}$$

$P(A=1|B)$

$= 1 - P(A=0|B)$

$P(A|B=1)$

$\neq 1 - P(A|B=0)$

WRONG

Joint Distribution:

Attr

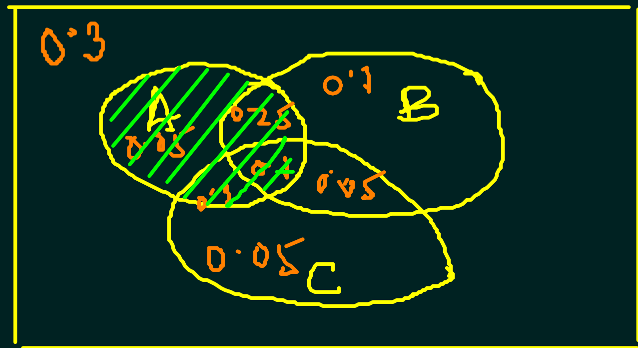
\# 3 attr

A	B	C	Prob
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10

TE

$$P(A=1) = \sum_{\substack{\text{rows} \\ \text{wher } A=1}} P(JDT) = 0{\cdot}5$$

$$P(A \wedge \bar{B}) = 0{\cdot}15$$



1

▷ Challenge: $(2^n - 1)$ est.

$\parallel$

▷ 100 attribute $\longrightarrow \approx 2^{100} \approx 10^{30}$

$\longrightarrow$ \# Population $\approx 10^9 / 10^{10}$

▷ Data Sparsity: 0·9999 (null)

$$P(W \mid \langle 100 \rangle) = \frac{\sum - \{0, 1\}}{\sum -} = ??$$

$f: \langle HW, G \rangle \rightarrow W$ → Poor / Rich

$\text{Prob}[W \mid HW, G]$ ← est.

gender	hours_worked	wealth	
Female	v0:40.5-	poor	0.253122
		rich	0.0245895
	v1:40.5+	poor	0.0421768
		rich	0.0116293
Male	v0:40.5-	poor	0.331313
		rich	0.0971295
	v1:40.5+	poor	0.134106
		rich	0.105933

$\text{Prob}(\text{Poor}) =$

$$\text{Prob}(\text{Male} \mid \text{Poor}) = \frac{\text{Prob}(M \wedge P)}{P(P)} = \frac{0.465}{0.\cancel{7}5}$$

✓ ✓

JPDT

$$\text{Prob}(W = \text{Rich} \mid HW, G) = \frac{\text{Prob}(R \wedge <40.5- \wedge F)}{P(40.5- \wedge F)} = ✓ \; ?? \; ✓$$

$HW \rightarrow <40.5-$, $G \rightarrow f$

Are We Done?? → JPDT availble ✓ (awesome!!)

MLE MAP → [Learn Smart Prob. Estimates]

[Represent JPDT efficiently] ← (Bayes Net)

▷ How to Est. Prob. Smartly?
(learn)

H (X = 1) T (X = 0)
Coinflip

→ α_1 heads + α_0 tails

$$\hat{\theta} = P(X=1) = \boxed{\frac{\alpha_1}{\alpha_1 + \alpha_0}}$$

MLE

▷ 49 Heads + 51 Tails
{Train Data Abund.

$$P(X=1) = \frac{49}{100} = 0.49$$

▷ 2 Heads + 1 Tails
{Less Train Data

$$P(X=1) = \frac{2}{3} = 0.67$$

??

→ "A PRIOR" knowledge → 0.5 → $P(\theta)$

10_H, 10_T } 20

[Online Learning Algo] { If data is less I am biased to my Prior
If data is more I follow data likelihood

MAP → $$\hat{\theta} = P(X=1) = \frac{(\alpha_1 + 10)}{(\alpha_1 + 10) + (\alpha_0 + 10)}$$

100000

$\boxed{\text{MLE}}$: $\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \left(\text{Prob}(\text{Data} \mid \theta) \right)$

$\boxed{\text{MAP}}$: $\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} \left(\text{Prob}(\theta \mid \text{Data}) \right)$

$$\frac{P(\text{Data} \mid \theta)\, P(\theta)}{P(\text{Data})}$$

likelihood

Prior

marginal

✓ $P(X=1) = \theta$

✓ $P(X=0) = 1-\theta$

$\boxed{1\ 0\ 0\ 1\ 0}$

$$\theta(1-\theta)(1-\theta)\theta(1-\theta) = \theta^2(1-\theta)^3$$

$$\frac{\partial}{\partial\theta}[\ ] = 0$$

$$\alpha_1(H) + \alpha_0 T \leftarrow \left[\theta^{\alpha_1}(1-\theta)^{\alpha_0}\right]$$

$$\frac{\partial}{\partial\theta}\left[\alpha_1 \ln\theta + \alpha_0 \ln(1-\theta)\right] = 0$$

$$\boxed{\frac{\partial \ln\theta}{\partial\theta} = \frac{1}{\theta}}$$

$$\Rightarrow \alpha_1 \frac{1}{\theta} + \alpha_0 \frac{d}{d\theta} \frac{\ln(1-\theta)}{1-\theta} \cdot (1-\theta) = 0$$

$$\Rightarrow \alpha_1 \frac{1}{\theta} + \frac{\alpha_0}{1-\theta}(-1) = 0$$

$$\Rightarrow \boxed{\hat{\theta}_{MLE} = \frac{\alpha_1}{\alpha_1 + \alpha_0}}$$