L1

- Information Theory & Coding

Diagram

Message source → Database
Message Receive → Label
} need strong connection
↓
maximize info

Context of ML

- Self-Information

Slide 2: when A occurs then, ⇒ lot of information
(if P(A) is low)

h(A) : qty of info

$$f(p^n) = f(p^{n-1} \cdot p) = n f(p)$$

$$f(p) = \left(f\left((p^{1/n})^n\right)\right) = n f(p^{1/n})$$

$$f(p^{m/n}) = f\left(\left(p^{\frac{1}{n}}\right)^m\right) = m f(p^{1/n})$$

$$\Rightarrow f(p^{1/n}) = \frac{f(p^{m/n})}{m} =$$

$$f(p^{1/n}) = \frac{f(p)}{n}$$

Exercise 1

$$h(B/A) = -\log_2 P(A,B)$$
$$= -\log_2 P(A) - \log_2 P(B/A)$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

when A and B are independent

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

$$\Rightarrow h(A,B) = -\log_2 P(A) - \log_2 P(B)$$
$$= h(A) + h(B)$$

Exercise 2

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

$$h(B/A) = -\log_2 P(A,B)$$
$$= -\log_2 P(A) - \log_2 P(B/A)$$

$$B \subset A \iff P(A/B) = 1$$

- $h(A,B) = h(A) + h(B/A)$

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)}$$ Bayes Theorem:

$$\Rightarrow P(B/A) = \frac{P(B)}{P(A)}$$

$$\Rightarrow h(A,B) = h(A) + h(B) - h(A)$$
$$= h(B)$$

$$h(A,B) = h(A) = h(B)$$

($h(A,B) \to$ whole diagram)

$$i(A,B) = h(A) - h(A/B) = h(B) - h(B/A)$$

$$A \subset B \Rightarrow p(B/A) = 1$$

$$\Rightarrow h(B/A) = -\log_2(B/A) = 0$$

$$i(A,B) = h(B) - \underbrace{h(B/A)}_{0}$$

$$= h(B)$$

(i) $i(A,B) = h(A,B) - h(B/A) - h(A/B)$

$= h(A) + h(B) \quad h(B) - h(A)$

$= 0$

(ii) $i(A,B) = h(A,B) - h(B/A) - h(A/B)$

$= h(B) - h(B/A) -$

- $h(A) = -\log P(A)$ $\quad P(A,B) = P(A)\,P(B/A)$
- $h(A,B) = -\log P(A,B) = -\log P(A) - \log P(B/A)$
- $h(B/A) = -\log P(B/A)$
- $h(B/A) = h(A,B) - h(A)$ $\quad \big|\ h(A,B) = h(A) + h(B/A)$

- $B \subset A \implies P(A/B) = 1$
- $A = B \implies A \subset B$ & $B \subset A$

- $i(A,B) = h(A) - h(A/B) = h(B) - h(B/A)$

- $H(S) = E[h(S)] = -\sum_{i=1}^{n} p_i \log p_i$ $\implies$ richness of source (entropy)
- $H(X) = -\sum_{i=1}^{n} P(X = x_i) \log P(X = x_i)$
- $H_2(p) = -p \log p - (1-p)\log(1-p)$
- $\lim_{p \to 0} p \log p = 0$

- Gibbs' inequality: $\sum_{i=1}^{n} p_i \log \frac{q_i}{p_i} \leq 0$ (p_i & q_i are probs)

- $H_n(p_1, \ldots, p_n) \leq \log n$ (entropy of n states is bound by log of n° of states)

$\downarrow$

maximum entropy obtained when $p_i = 1/n$ (same prob for all states)

- $H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j)$

$\downarrow$

$\to$ mn states $\implies$ bounded by $\log mn$

$\to$ symmetric

- $H(X,Y) = H(X) + H(Y)$ if X and Y are ind

- $H(X/Y=y_j) = -\sum_{i=1}^{n} P(X=x_i/Y=y_j)\log P(X=x_i/Y=y_j)$
  (y_j is fixed)

- $H(X/Y) = \sum_{j=1}^{m} P(Y=y_j)\, H(X/Y=y_j)$

- $H(X,Y) = H(X) + H(Y/X)$
  $= H(Y) + H(X/Y)$

- $H(X_1,\dots,X_n) = \sum_{i=1}^{n} H(X_i/X_1,\dots,X_{i-1})$

- $H(X) \leq H(X,Y)$ (since entropy is always +ve)
  $H(Y) \leq H(X,Y)$

- $H(X/Y) \leq H(X)$ (not proved)

- $I(X,Y) = H(X) - H(X/Y)$

$$= \sum\sum P(X=x_i, Y=y_j)\log\frac{P(X=x_i, Y=y_j)}{P(X=x_i)\,P(Y=y_j)}$$

Slide 23: Say when each one becomes equality
↓
$\leq$ inequalities

X, Y independent
or $X = f(Y)$ with f one-to-one application
(f^{-1})

depending on the inequality

If X & Y independent:

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X=x_i)\,P(Y=y_j)\log\left[P(X=x_i)\,P(Y=y_j)\right]$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X=x_i)\,P(Y=y_j)\left[\log P(X=x_i) + \log P(Y=y_j)\right]$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} P(X=x_i)\log P(X=x_i)\,P(Y=y_j) + P(Y=y_j)\log P(Y=y_j)\,P(X=x_i)$$

$$= -\sum_{i=1}^{n} P(X=x_i)\log P(X=x_i)\sum_{j=1}^{m} P(Y=y_j) - \sum_{j=1}^{m} P(y=y_i)\log P(Y=y_i)\sum_{i=1}^{n} P(X=x_i)$$

$$= -\sum_{i=1}^{n} P(X=x_i)\log\left[P(X=x_i)\right] + -\sum_{j=1}^{m} P(Y=y_i)\log P(Y=y_i)$$

$$= H(X) + H(Y)$$

L2 : Part 1 Exercises (Quantitative Measure of info)

Exercise 1

Event B : { Today is my birthday }

(i) $P(B) = 1/365$

$h(B) = -\log_2 P(B) = -\log_2 (1/365) = \log_2 (365) = 8.5$

(ii) $H(s_i) = -\sum_{i=1}^{2} p_i \log p_i = \frac{1}{365} \log_2 365 + \frac{364}{365} \log_2 \left(\frac{365}{364}\right)$

$= 0.028$

Exercise 2

64 32 16 8 4 2 1

$P(x_i) = 1/64$

$h(x_i) = -\log_2 1/64 = \log_2 (64)$

$= 6$

$H[x_i] = -\sum_{i=1}^{64} \frac{1}{64} \log_2 \left(\frac{1}{64}\right) = 6$

$H(S) = \log_2 (64) = 6$ sh/state

= number of question

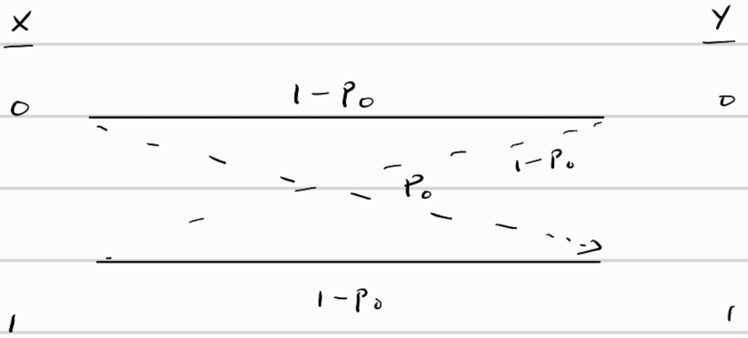
Interpretation: Each question → 1 sh (2 responses per each question since same prob for each response)

⇒ We require 6 sh of info to find the piece

⇒ 6 questions needed

⇒ That's why we cut board in half (2 equal parts) → max info each time (1 sh)

Ex 6



$P(X=0) = 1/4$

$P(X=1) = 3/4$

$P_0 = P(Y=0 \mid X=1)$

$= P(Y=1 \mid X=0) = 0.1$

$1-P_0 = P(Y=0 \mid X=0)$

$= P(Y=1 \mid X=1)$

$= 0.9$

• $H(X) = -\sum_{i=0}^{1} P(X=i) \log_2 P(X=i)$

$= -\left[\frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{3}{4}\log_2\left(\frac{3}{4}\right)\right]$

$= -\left[\frac{1}{4}(-2) + \frac{3}{4}(-0.415)\right]$

$= 0.811$ sh/state of X

$(Y=0) = (Y=0 \ \& \ X=1)$ or $(Y=0 \ \& \ X=0)$

• $P(Y=0) = P(X=0)(1-P_0) + P(X=1)P_0$

$= 0.25(0.9) + 0.75(0.1)$

$= \frac{3}{10}$

$P(Y=1) = P(X=0)P_0 + P(X=1)(1-P_0)$

$= (0.25)(0.1) + 0.75(0.9)$

$= 0.025 + 0.675$

$= 0.7 \quad (1 - 3/10)$

$H[Y] = -\left[3/10 \log_2 3/10 + 0.7 \log_2(0.7)\right]$

$=$

$$\cdot\ H(X,Y) = -\sum_{i=0}^{1}\sum_{j=0}^{1} P(X=x_i, Y=y_i)\log_2 P(X=x_i, Y=y_i)$$

$$P(X=0, Y=0) = 9/40 \qquad P(X=0)\ P(Y/X=0)$$

$$P(X=1, Y=0) = 7/10 \qquad P(X=1)\ P(Y/X=1)$$

$$P(X=0, Y=1) = 0.025 \qquad P(X=0$$

$$P(X=1, Y=1) = 0.675$$

$$H(X/Y=y_j) = -\sum_{i=1}^{n} P(X=x_i/Y=y_j)\log P(X=x_i/Y=y_j)$$

(y_j is fixed)

$$H(X/Y) = \sum_{j=1}^{n} P(Y=y_j)\, H(X/Y=y_j)$$

$$H(Y/X) = \sum_{i=1}^{n} P(X=x_i)\, H(Y/X=x_i)$$

$$\rightarrow H(Y/X=0) = -P(Y=0/X=0)\log P(Y=0/X=0)$$

$$-P(Y=1/X=0)\log P(Y=1/X=0)$$

0.3322

0.1368

$$= -(0.9\log_2(0.9)) - (0.1\log(0.1))$$

$$= 0.469$$

$$H(Y/X=1) = -P(Y=0/X=1)\log P(Y=0/X=1)$$

$$-P(Y=1/X=1)\log P(Y=1/X=1)$$

$$= -0.1\log 0.1 + 0.9\log 0.9$$

$$= 0.469$$

$H(Y/X) = 0.25(0.469) + 0.75(0.469)$

$= 0.469$

$H(X,Y) = H(X/Y) + H(Y)$

$= H(Y/X) + H(X)$

$\Rightarrow H(Y/X) = 0.4$

$H(X/Y) = 0.47$

$I(X,Y) = H(X,Y) - H(X/Y) - H(Y/X)$

Ex 3

$X \sim Geo(1/2)$

$P(X=k) = 0.5^{k-1} \, 0.5 = 0.5^{k}$

$H(X) = -\sum_{n=1}^{\infty} 0.5^{n} \log(0.5)^{n}$ $\quad n \log 0.5$

$= n \log(2)^{-1}$

$= - \sum 0.5^{n} . n \log(1/2)$ $\quad = -n$

$= \sum_{n=1}^{\infty} n \, 0.5^{n}$

$= \frac{0.5}{(1-0.5)^{2}} = \frac{0.5}{0.25} = 2$

Problem 2

1) 18 hypothesis — 1st coin heavier than others
since 9 coins — lighter
2nd coin ...

$$H(S) = \log_2 18 \text{ sh / scenario}$$

2) 3 possible outcomes : Equilibrium (P_e)
Left heavier (P_L)
Right heavier (P_R)

$$H_{max}(P) = \log_2 3 \text{ sh/weight (experiment)}$$

Min n° of operations $= \dfrac{\log_2 18}{\log_2 3} \simeq 2.63$ experiments

3) $$P_{eq} = \frac{{}^{2n}C_8}{{}^{2n}C_9} = \frac{\dfrac{8!}{(8-2n)!\,2n!}}{\dfrac{9!}{(9-2n)!\,2n!}} = \frac{9-2n}{9} = 1 - \frac{2n}{9}$$

$$P_L = P_r = \frac{1}{2}\left(1 - \left(1 - \frac{2n}{9}\right)\right) = \frac{n}{9}$$

4) Cond: to get max of info

$$P_e = P_r = P_L \iff 1 - \frac{2n}{9} = \frac{n}{9} \Rightarrow \underline{n = 3}$$

Part 2 : Exercises

1) $H_{max}(X) = \log_2(m)$

When X is uniformly distributed

2) $H(Y) = \log_2(n)$

3) $0 \le H(X/Y) \le H(X) \le H(Y) \le H(X,Y) \le H(X)+H(Y)$

4) Equality: X & Y are independent

Equality if X is uniformly dist

Equality if $X \subset Y$

Equality if X & Y are independent

Cedric.richard@unice.fr

5) $P(Y=y_1) = 1/6 + 1/24 + 1/24 = 1/4$

$P(Y=y_2) = \frac{1}{12} + \frac{1}{8} + \frac{1}{24} = \frac{1}{4}$

$P(Y=y_3) = 1/4$

$P(Y=y_4) = 1/4$

6) $I(X,Y) = H(Y) - H(Y/X)$

$= 2 -$

$H(Y/X) = \sum_{i=1}^{3} H(Y/X=x_i)\, P(X=x_i)$

To calculate $H(Y/X=x_i)$ we need the dist of $P(Y/X=x_i)$

Use $P(Y = y_j \mid X = x_i) = \dfrac{P(X = x_i, Y = y_i)}{P(X = x_i)}$

$P(Y/X)$	Y_1	Y_2	Y_3	Y_4
x_1	$1/8$	$1/4$	$1/2$	$1/8$
x_2	$1/3$	$1/4$	$1/12$	$1/3$
x_3	$1/4$	$1/4$	$1/4$	$1/4$

we divided the prob in the table by $P(X = x_1) = 1/3$

$$H(Y / X = x_1) = H(1/8, 1/4, 1/2, 1/8)$$

$$H(Y/X = x_2) = \ldots$$

$$H(Y/X = x_3) = \ldots$$

$$\therefore I(X, Y) = 2 - \Big[\tfrac{1}{3} H(1/8, 1/4, 1/2, 1/8) + \tfrac{1}{2} H(\ldots) + \tfrac{1}{6} H(\ldots) \Big]$$

$$2 - \left(\frac{1.75}{3} + \frac{1.85}{2} + \frac{2}{6} \right)$$

$\frac{1.85}{2}$ 0.92

$$= 2 - 0.58 - 0.925 + 0.33$$

$$= 0.15$$

Problem

1) $P(M=0) = 3/16 + 1/8 = 5/16$

$P(M=1) = 5/8 + 1/16 = 11/16$

$P(T=1) = 5/8 + 3/16 = 13/16$

$P(T=0) = 3/16$

$P_{err}(M) =$

2) $P(M=1, T=0) + P(M=0, T=1)$

$= 1/16 + 3/16 = 4/16 = 1/4$

3) $P_{err}(S) = P(S=1, T=0) + P(S=0, T=1)$

$= 3/16$

$\underline{3/16} < 4/16$

<u>Student is better</u>

4) $I(E,T) = H(E) - H(E/T)$

$= 0$

5) $I(T,M) = H(T) + H(M) - H(T,M) \approx 0.091$

7) $P(E=1) = 496/512 = 31/32$

$P(\bar{T}=0) = 16/512 = \frac{1}{32}$