ORIE 4742 - Info Theory and Bayesian ML

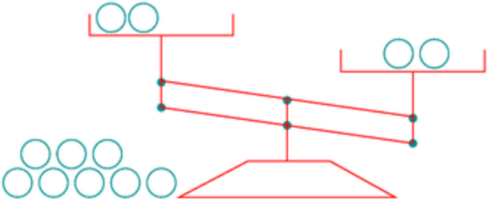
Lecture 3: Measuring Information

February 15, 2021

Sid Banerjee, ORIE, Cornell

Mackay’s weighing puzzle

The weighing problem



You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

Design a strategy to determine

which is the odd ball

and whether it is heavier or lighter,

in as few uses of the balance as possible.

how much 'information' does a random variable have?

- 2 state lotteries S_1, S_2, winning number is $X_1 = 1$, $X_2 = 1$

Suppose $S_1 \equiv$ Vermont, $S_2 \equiv$ Texas (N_1 = # of people in lottery $1 \ll N_2$)

- If we do not know X_1, X_2, then is $X_1 = 1$ or $X_2 = 1$ more surprising?

- Is $X_1 = 1$ more/less surprising than $X_1 = 12793$

axioms of 'information' - info exists only if uncertainty
- the exact information does not matter (only the 'surprise' matters)
- more 'surprising' r.v. have more info

(Shannon '48)

Idea - Information of a rv $\equiv$ amount of uncertainty resolved by knowing the r.v.

reading assignment: chapter 4 of Mackay

quantifying information content

measuring information

consider (discrete) rv X taking values $\mathcal{X} = \{a_1, a_2, \ldots, a_k\}$, with probability mass function $\mathbb{P}[X = a_i] = p_i \,\forall\, i, \sum_{i=1}^{k} p_i = 1$

Shannon's entropy function

- outcome $X = a_i$ has *information content*

$$h(a_i) = \log_2\left(\frac{1}{p_i}\right)$$

p_i large $\Rightarrow$ $h(a_i)$ is small

← fn of a_i but does not depend on a_i

'convention' (bits)

- random variable X has *entropy*

$$H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right)$$

x	$h(x)$	$p(x)$
a_1	$\log_2(1/p_1)$	p_1
a_2	$\log_2(1/p_2)$	p_2
$\vdots$	$\vdots$	$\vdots$
a_k	$\log_2(1/p_k)$	p_k

entropy: basic properties

Shannon's entropy function

- outcome $X = a_i$ has *information content*: $h(a_i) = \log_2\left(\frac{1}{p_i}\right)$
- random variable X has *entropy*: $H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right)$

- only depends on distribution of X (i.e., $H(X) = H(p_1, p_2, \ldots, p_k)$)
- $H(X) \geq 0$ for all X $\quad$ ($\because \log(1/p_i) \geqslant 0 \ \forall i$)
- if $X \perp\!\!\!\perp Y$, then $H(X, Y) = H(X) + H(Y)$

  where joint entropy $H(X, Y) \triangleq \sum_{(x,y)} p(x, y) \log_2 1/p(x, y)$

independent

$\not\perp\!\!\!\perp$: not indep

$$= \sum_{(x,y)} p(x)p(y)\left(-\log_2 p(x) - \log_2 p(y)\right)$$

$$= \left(\sum_x -p(x)\log p(x)\right) + \left(\sum_y -p(y)\log p(y)\right)$$

entropy: basic properties

Shannon's entropy function

- outcome $X = a_i$ has *information content*: $h(a_i) = \log_2\left(\frac{1}{p_i}\right)$
- random variable X has *entropy*: $H(X) = \mathbb{E}[h(X)] = \sum_{i=1}^{k} p_i \log_2\left(\frac{1}{p_i}\right)$

- if $X \sim$ uniform on $\mathcal{X}$, then $H(X) = \log_2|\mathcal{X}|$; else, $H(X) \leq \log_2|\mathcal{X}|$

(1) ($|\mathcal{X}| = k$) (2)

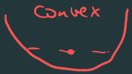
(1) $-\sum_{i=1}^{|\mathcal{X}|} p_i \log p_i = -\sum_{i=1}^{|\mathcal{X}|} \frac{1}{|\mathcal{X}|} \log \frac{1}{|\mathcal{X}|} = \log|\mathcal{X}|$

(2) $\forall \{p_i\}$ s.t. $p_i \geq 0$, $\sum_{i=1}^{|\mathcal{X}|} p_i = 1$, $\max -\sum_{i=1}^{|\mathcal{X}|} p_i \log p_i \leq \log_2|\mathcal{X}|$

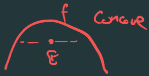
Idea – $H(X) = E[h(X)]$ where $h(x) = -\log p(x)$

- $\mathbb{E}[h(x)] = \mathbb{E}[\log_2(1/p(x))]$

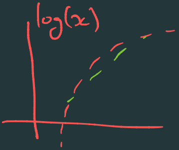
- (Jensen's) $\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$



$\leq f(\mathbb{E}[x])$



$\Rightarrow \mathbb{E}[\log(g(x))] \leq \log_2(\mathbb{E}[g(x)])$



$\Rightarrow \mathbb{E}[h(x)] = \mathbb{E}[\log_2(1/p(x))]$

$\leq \log_2[\underbrace{\mathbb{E}[1/p(x)]}_{\sum_{i=1}^{|x|} p_i \cdot (1/p_i) = |x|}]$

$= \log_2|x|$

designing questions to maximize information gain (heuristic)

the game of 'sixty three'

guess number $X \in \{0, 1, 2, \ldots, 62, 63\}$

Q: how many (and what) Yes/No questions should you ask?

Ideal:

Q1 - Is $X \geq 32$? — Yes → Is $X \geq 16$ … ; No → …

Binary search

# of questions $= 6 \geq H(X)$ ($H(X) = 6$ if $X \sim \mathrm{Unif}\{0, \ldots, 63\}$)

Q1 - Is X even? — Yes → Is $X/2$ odd or even? … ; No → Is $X+1/2$ odd or even? …

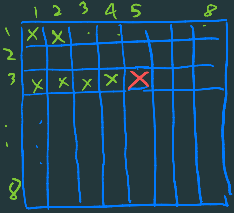
Claim - Amount of entropy in each answer $= 1$ bit

designing questions to maximize information gain

the game of 'submarine'

player 1 hides a submarine in one square of an 8×8 grid

player 2 shoots at one square per round



- $\mathcal{X} = \{(x,y) ; x \in \{1,\dots,8\}, y \in \{1,\dots,8\}\}$

  If $X \sim \mathrm{Unif}(\mathcal{X})$, then $H(X) = 6$ $(= 3+3)$ ← $\mathrm{Unif}(\{1,\dots,8\})$

- Question $\equiv (Q_x, Q_y)$

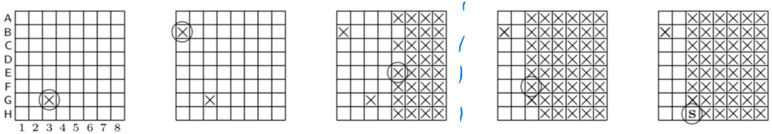
- $Q_1 \equiv$ Is $(x,y) = (1,1)$? $\quad h(Y_1) = -\frac{1}{64}\log_2\frac{1}{64} - \frac{63}{64}\log\frac{63}{64}$
- If $Y_1 = \text{No}$, $Q2 \equiv$ Is $(x,y) = (1,2)$? $\quad h(Y_2) = -\frac{1}{63}\log_2\frac{1}{63} - \frac{62}{63}\log_2\frac{62}{63}$

designing questions to maximize information gain

the game of 'submarine'

player 1 hides a submarine in one square of an 8×8 grid

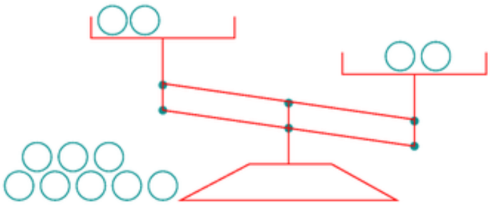
player 2 shoots at one square per round



move #	1	2	32	48	49
question	G3	B1	E5	F3	H3
outcome	$x = \mathtt{n}$	$x = \mathtt{n}$	$x = \mathtt{n}$	$x = \mathtt{n}$	$x = \mathtt{y}$
$P(x)$	$\frac{63}{64}$	$\frac{62}{63}$	$\frac{32}{33}$	$\frac{16}{17}$	$\frac{1}{16}$
$h(x)$	0.0227	0.0230	0.0443	0.0874	4.0
Total info.	0.0227	0.0458	1.0	2.0	6.0

Mackay's weighing puzzle

The weighing problem



You are given 12 balls, all equal in weight except for one that is either heavier or lighter.

Design a strategy to determine

which is the odd ball

and whether it is heavier or lighter,

in as few uses of the balance as possible.

information acquisition in the weighing puzzle

What is the best you can do? $X \equiv$ true outcome

- $\mathcal{X} \equiv$ set of outcomes $= \{(1,h),(1,l),(2,h),(2,l) \ldots (12,l)\}$

$\Rightarrow |\mathcal{X}| = 24 \Rightarrow H(X) = \log_3 24$ trits $= \log_2 24$ bits

- Consider each weighing - 3 outcomes - LH, RH, E

max info per weighing $= \log_3 3 = 1$ in 'trits'

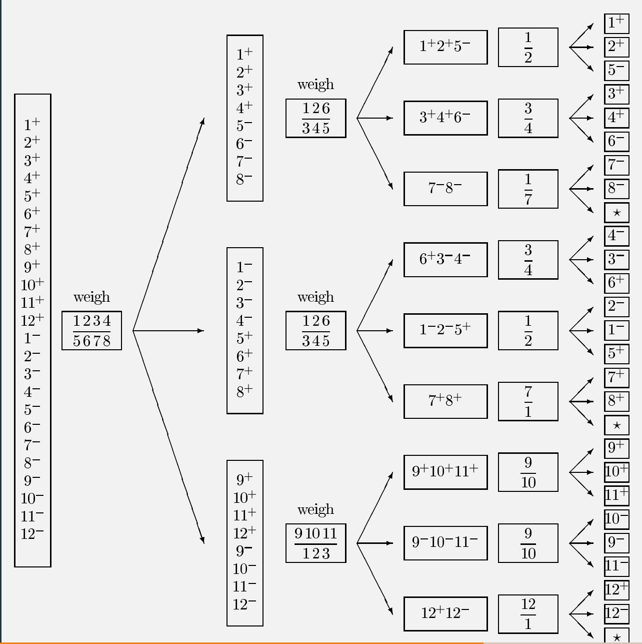
(or $\log_2 3$ bits)

$\Rightarrow$ Need k questions s.t. $k \log_2 3 \geq \log_2 24$

$\Rightarrow k \geq 3$

information acquisition in the weighing puzzle

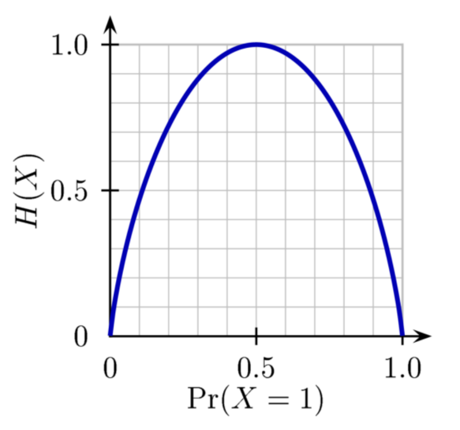
weighing game: an optimal solution



Reading
- Ch 2 of Mackay
- Ch 1, Ch4 ≡ Source coding

binary entropy function

if X Bernoulli(p), then $H(X) \triangleq H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$



- (useful formula) for any $k, N \in \mathbb{N}$, $k \leq N$: $\binom{N}{k} \approx 2^{NH_2(k/N)}$

conditional entropy

suppose $X \sim \{p_1, p_2, p_3, p_4\}$, and let $Y = \mathbb{1}_{[X \in \{a_1, a_2\}]}$; then we have

$$H(X) = H(Y) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}\right) + (p_3 + p_4) H_2\left(\frac{p_3}{p_3 + p_4}\right)$$

conditional entropy

suppose $X \sim \{p_1, p_2, p_3, p_4\}$, and let $Y = \mathbb{1}_{[X \in \{a_1, a_2\}]}$; then we have

$$H(X) = H(Y) + (p_1 + p_2) H_2\left(\frac{p_1}{p_1 + p_2}\right) + (p_3 + p_4) H_2\left(\frac{p_3}{p_3 + p_4}\right)$$

conditional entropy

for any rvs X, Y: $H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y)$
$= \sum_{y \in \mathcal{Y}} p(y) \sum_{x \in \mathcal{X}} p(x|y) \log_2(1/p(x|y))$

the chain rule

the chain rule (information content)

for any rvs X, Y and realizations x, y:

$$h(x, y) = h(x) + h(y|x) = h(y) + h(x|y)$$

the chain rule

the chain rule (entropy)

for any rvs X, Y:

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

blank

blank

blank