ORIE 4742 - Info Theory and Bayesian ML

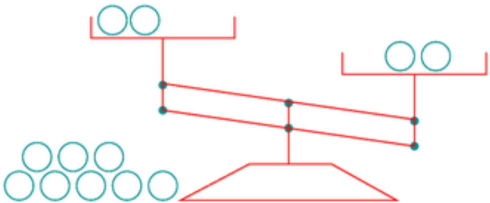
Lecture 3: Information Measures and Data Compression

January 27, 2020

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?

- X discrete r.v., $X \in \mathcal{X} = \{a_1, a_2, \dots, a_k\}$

  $\mathbb{P}[X = a_i] = p_i \quad \forall i$

- 'Information' in X $\equiv$ 'surprise' of seeing an outcome

- <u>Useful</u> - How many bits do I need to communicate to tell you $X = a_i$ (on average)

Eg. $X \sim \text{Unif}\{0\}$ - 0 bits

$X \sim \text{Unif}\{0,1\}$ - 1 bit

$X \sim \text{Unif}\{0,\dots,7\}$ - 3 bits

assuming both parties know p_i

reading assignment: chapter 4 of Mackay

quantifying information content

measuring information

consider (discrete) rv X taking values $\mathcal{X} = \{a_1, a_2, \dots, a_k\}$, 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*

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

- random variable X has *entropy*

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

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

  information hidden in X

- 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
- if $X \perp\!\!\!\perp Y$, then $H(X, Y) = H(X) + H(Y)$

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

Pf – $H(X,Y) = \sum_{(x,y)} -P_X(x)\,P_Y(y)\,\log_2\left(P_X(x)\,P_Y(y)\right)$

$$= -\sum_x \sum_y \left(P_X(x)\,P_Y(y)\,\log_2 P_X(x)\right) - \sum_y \sum_x P_X(x)P_Y(y)\log_2(P_Y(y))$$

$$= -\sum_x P_X(x)\,\log_2 P_X(x) \quad - \sum_y P_Y(y)\log P_Y(y)$$

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, $H(X) \leq \log_2 |\mathcal{X}|$

. If $p_i = 1/|\mathcal{X}|$ $\forall a_i \in \mathcal{X}$, then $\sum_i p_i \log_2 1/p_i = \sum_i \frac{1}{|\mathcal{X}|} \log_2 |\mathcal{X}| = \log_2 |\mathcal{X}|$

. $\forall p_i$, $\sum_{i=1}^{|\mathcal{X}|} p_i = 1$, $p_i \geq 0$, $H((p_1, \dots, p_{|\mathcal{X}|})) = \sum_{i=1}^{|\mathcal{X}|} p_i \log_2 \frac{1}{p_i} \leq \log_2 |\mathcal{X}|$

$$H(X) = \sum_{i=1}^{|\mathcal{X}|} p_i\, h(a_i) = \mathbb{E}[h(X)] = \mathbb{E}[\log_2(g(X))] \quad 1/p(x)$$

$$\underbrace{\leq}_{\text{Jensen's, since } \log(x) \text{ is concave}} \log_2 \mathbb{E}[g(X)] = \log_2\left(\sum_{i=1}^{|\mathcal{X}|} p_i \left(\frac{1}{p_i}\right)\right) = \log_2 |\mathcal{X}|$$

designing questions to maximize information gain

the game of 'sixty three'

guess number $X \in \{0, 1, 2, \dots, 62, 63\}$, Assume $X \sim \mathrm{Unif}(\{0,\dots,63\})$

- $\underline{Q1}$ - Is $X \geq 32$

  Y_1:
  - 1: $X \in \{32, 33, \dots, 63\}$ wp $1/2$
  - 0: $X \in \{0, 1, \dots, 31\}$ wp $1/2$

  $$h(Y_1 = 1) = h(Y_1 = 0) = 1 \text{ bit}$$

- $\underline{Q2}$ - If $Y_1 = 0$, Is $X \geq 16$

  Y_2:
  - 1: $X \in \{16, \dots, 31\}$ wp $1/2$
  - 0: $X \in \{0, \dots, 15\}$ wp $1/2$

  $$h(Y_2 = 1 \mid Y_1) = h(Y_2 = 0 \mid Y_1) = 1 \text{ bit}$$

$\vdots$

- Need 6 questions, each gives 1 bit of info content
- Note $6 = \log_2 64 = \log_2 |\{0, \dots, 63\}| = H(X)$

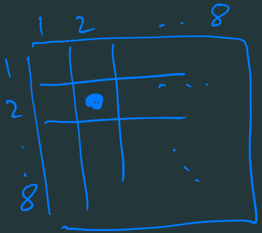
designing questions to maximize information gain

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

X = posn of submarine



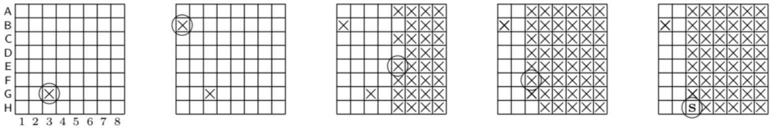
- Set of questions = (x,y), $x,y \in \{1,\dots,8\}$
- $H(X) \leq \log 8 \times 8 = 6$

designing questions to maximize information gain

the game of 'submarine'

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

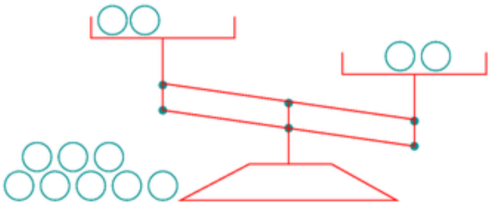
player 2 shoots at one square per round



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?

index of odd ball

- $\mathcal{X}$ = set of all universes = $\{(1,H),(2,H),\dots,(12,H),(1,L),(2,L),\dots,(12,L)\}$

weight of odd ball

$$\Rightarrow H(X) = \log_2(|X|) = \log_2 24 \text{ bits } \left(\begin{smallmatrix}\text{assuming}\\ \text{uniform}\end{smallmatrix}\right)$$

- Each question has 3 outcomes - Left heavier (L), Right heavier (R), Equal (E)

$$\Rightarrow H(Q_i) \leq \log_2(3) \text{ for each response } Q_i$$

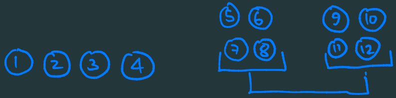
- Thus # of questions required $\geq \left\lceil \frac{\log_2 24}{\log_2 3} \right\rceil = \frac{\log_2 27}{\log_2 3} = \underline{\underline{3}}$

information acquisition in the weighing puzzle

How to design questions?

<u>Heuristic</u> - Choose Q_i to maximize information gain

- <u>For Q_1</u> -



$P_L = P_R = P_E = 1/3$

$\Rightarrow H(Q_1) = \log_2 3 \approx 1.58$

ALT -

$P_L = P_R = 5/12, \quad P_E = 2/12$

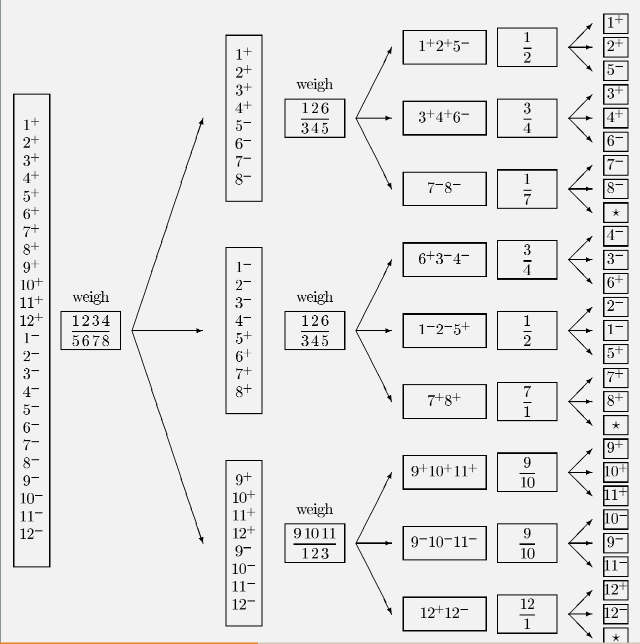
$\Rightarrow H(Q_1) = \frac{5}{6}\log_2\frac{12}{5} + \frac{1}{6}\log_2\frac{12}{2} = 1.48$

$\leq$

- Can choose Q_2, Q_3 similarly to ensure L, R, E are close to uniform.

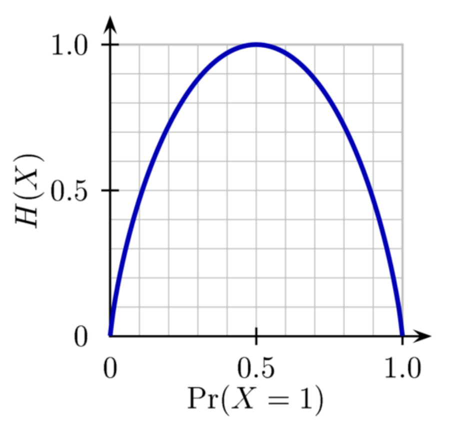
- <u>Note</u> - Just maximizing $H(Q_i \mid Q_1, Q_2, \ldots, Q_{i-1})$ is not sufficient; however it is a very good rule of thumb.

weighing game: an optimal solution



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

conditional entropy

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

for any rvs X, Y:

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