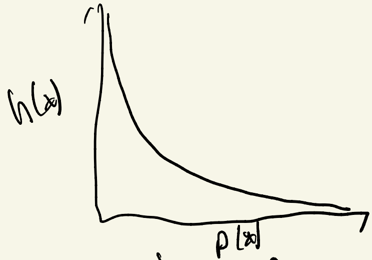
1. Intro to entropy & mutual information

2. Source coding ← weighing / Langlin eg.

3. Channel coding ← typewriter



① Shannon surprise (or info content) of an event

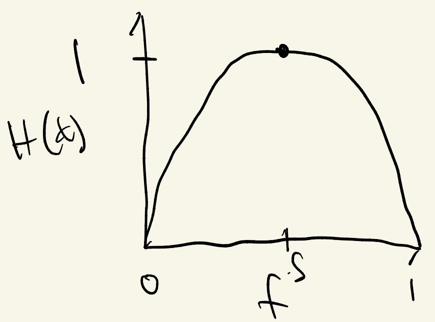
$x \in X$

$$h(x) = \log_2 \frac{1}{p(x)}$$

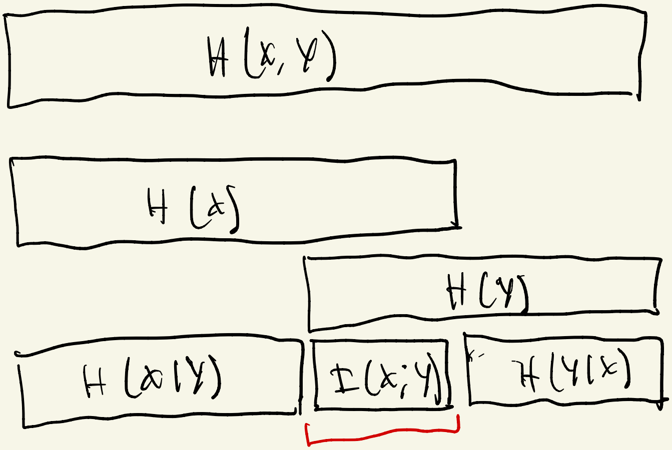
entropy of a R.V. is the average surprise across all events

$$H(X) = \sum_{x \in X} p(x) \log \frac{1}{p(x)} = -\sum_{x \in X} p(x) \log p(x)$$

flipping a bent coin $x = \begin{cases} \text{heads} & p = f \\ \text{tails} & p = 1-f \end{cases}$



uniform prob maximizes entropy for more outcomes too now, for two R.V.s X ∩ Y

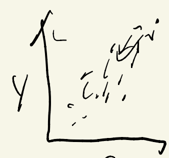


$$H(X,Y) = -\sum_{x \in X, y \in Y} p(x,y) \log p(x,y)$$

$$H(Y|X) = \sum_{x \in X} p(x)\, H(Y \mid X = x)$$

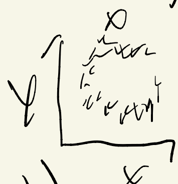
$$I(X;Y) \overset{?}{=} H(X) - H(X|Y) = H(Y) - H(Y|X)$$

$$= \sum_{x \in X, y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$



$r > 0$

$I > 0$



$r = 0$

$I \geq 0$

Kullback Leibler divergence

$$I(X;Y) = D_{KL}\left(p(x,y) \,\|\, p(x)\,p(y)\right)$$

$$D_{KL}(P \,\|\, Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$

② source coding : how many bits do we need to convey a given piece of information

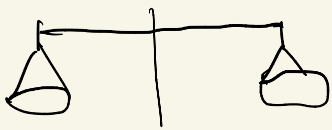
how far can we compress an image/text w/o losing information?

<u>exercise</u>

you have 12 balls, all are equal weight except 1

this ball is either heavier or lighter

you have a scale with two pans, any number of balls can be put in each pan



∧ each outcome can be left heavier, right heavier, or equal

<u>task</u> : identify <u>which</u> ball is different ∧ if <u>heavier</u> or <u>lighter</u>

w/ as few uses of the balance as possible

5 min <u>tip 1</u>: build a tree for your strategy

5 min <u>tip 2</u>: how many of you are weighing 6 x 6 balls? 4 x 4?

What is the information content of each?

S min — Minimum number of weighings?

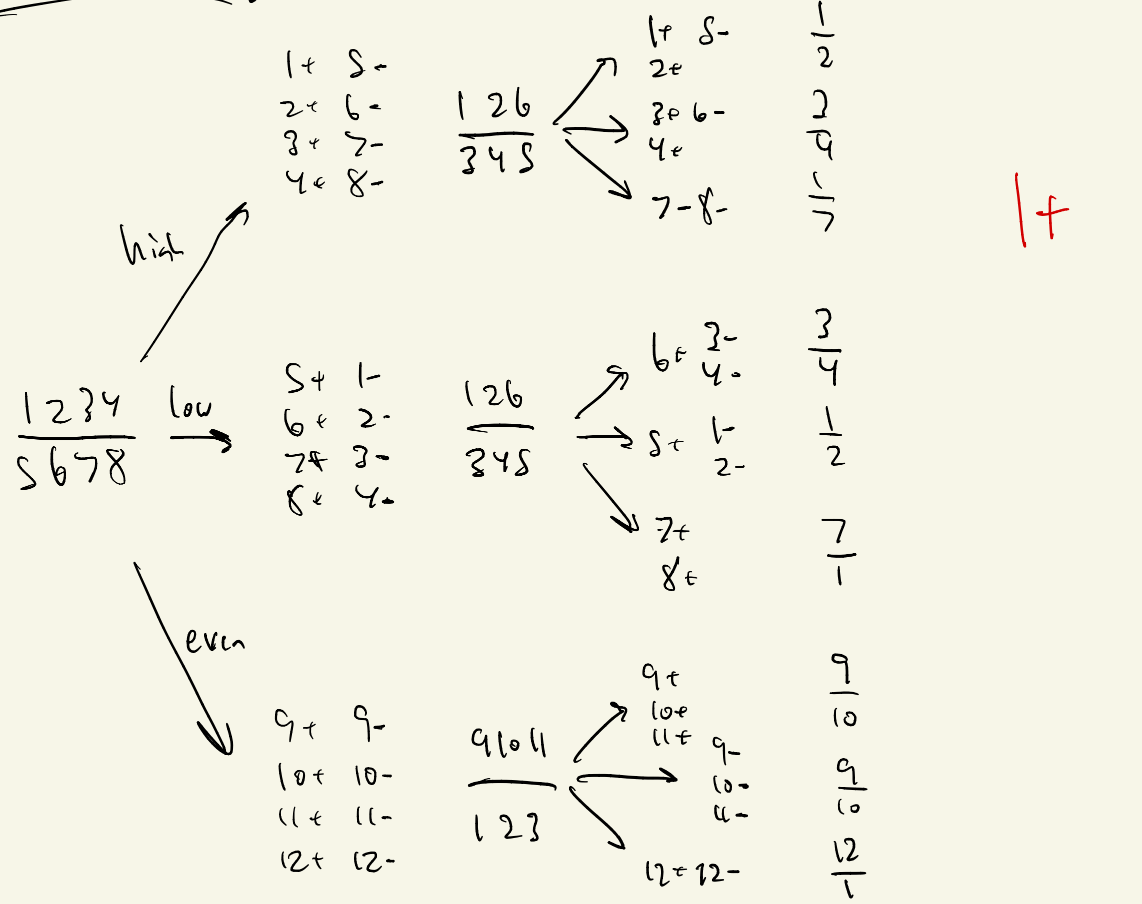
Inf. theory tells us how many weighings you need

24 possible outcomes → $\log 24$ bits of info

each weighing can give us $\log 3$ bits

→ so, $3 \log 3 = \log 27$ → minimum 3 weighings

an optimal strategy



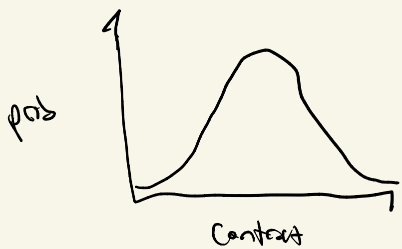
two principles : (a) maximize entropy for outcome w/i coding

(b) minimize redundancy across coding

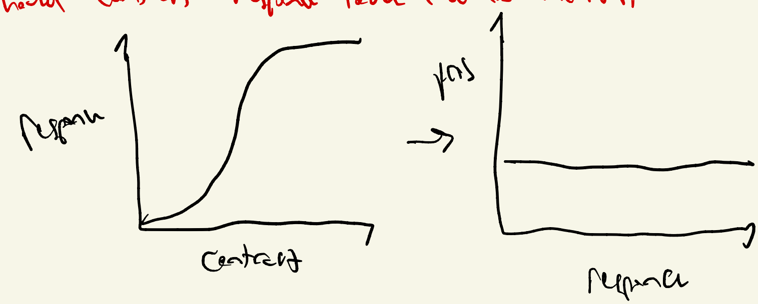
how can we apply this to the neural code?

(a) contrast sensitivity in fly compound eye

if distribution of contrast in natural scene is:

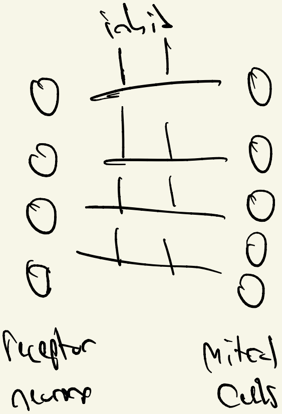


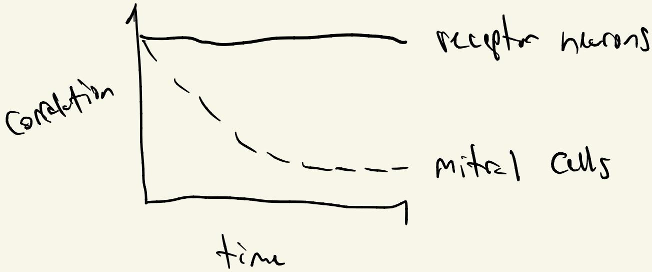
What should contrast-response look like in neuron?



Laughlin, 1981

(b) odor representations in olfactory bulb





need fewer neurons to distinguish odors

Friedrich & Laurent, 2001

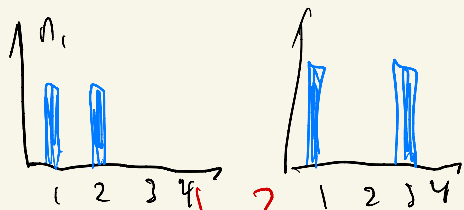
redundancy betw. diff neurons is minimized

designing a neural code for numbers

1, 2, 3, 4

what is tuning of one neuron?

okay, what should tuning of second neuron be?



is this what you expect a neural code to look like?

why not?

noise! (but also other stuff) sparse coding

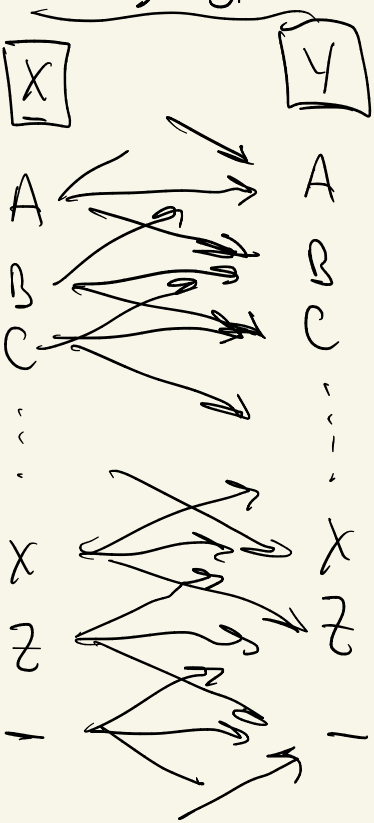
Olshausen & Field, 1996

② channel coding

$X \sim x \to \boxed{\text{channel}} \to y \sim Y$

how can we convey information reliably in the presence of noise?

the noisy typewriter



$P(x \neq y) = \frac{2}{3} = P_E = 1 - P_C$

how can we do better than this?

Repetition (R_2) code

$A_X = \{AA, BB, \ldots, \_\_\}$

$P_X = \{\frac{1}{27}, \ldots, \frac{1}{27}\}$

$X = BB$

$y \in \{\underbrace{AA, BB, CC,}\ \underbrace{CA, AC,}\ \underbrace{BC, CB, AB, BA}\}$

$P_C = \frac{1}{9} \cdot \frac{1}{3} \cdot 3 + \frac{1}{9} \cdot 1 \cdot 2 + \frac{1}{9} \cdot \frac{1}{2} \cdot 4$

$= \frac{5}{9} \to P_E = 1 - \frac{5}{9} = \frac{4}{9} < \frac{2}{3}$

Non-confusible subset code

$A_x^1 = \{B, E, \dots, Z\}$ → $A_x^2 = \{BB, BE, \dots, ZZ\}$

$P_x^1 = \{\frac{1}{9}, \dots, \frac{1}{9}\}$ → $P_x^2 = \{\frac{1}{81}, \dots, \frac{1}{81}\}$

$P_C = 1$, $P_E = 0$

is this as good as it gets?

The capacity of a channel

$$C = \max_{P_x} I(X;Y)$$

is the amt of info in bits that can be transmitted down channel w/o errors

given them 5 min to do

$$C_{NT} = \max_{P_x} H(Y) - H(Y|X)$$

talk through

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X=x)$$

p(x) doesn't matter

$$= H(Y|X=B)$$

$$= \sum_{y \in Y} p(y|X=B) \log \frac{1}{p(y|X=B)}$$

$$= \log 3$$

$$H(Y) = \log 27$$

p(x) being uniform maximizes this

$$C_{NT} = \log 27 - \log 3 = \log 9$$

for non-confusable subset:

$$I(X;Y) = H(X) - H(X|Y)$$
$$= \log 9 - 0$$
$$= \log 9$$

it conveys the max info possible w/o making errors!

two principles:

(a) transmission in noisy discrete systems can be done w/o errors at finite energy / neurons

(b) repetition / naive redundancy is not efficient for reducing errors

for neurons, error scales as $\frac{1}{N}$ w/ repetition

but $\frac{1}{N^2}$ or even $\exp(-N)$

w/ other approaches

Kim, Schwab, Fiete (2020)

Sreenivasan & Fiete (2011)