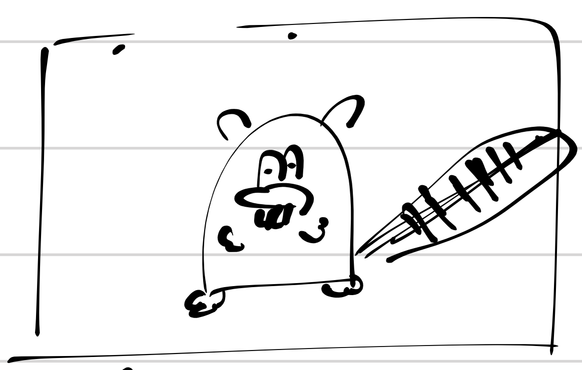
lectures: online
classes: offline
→ wiki page of the course

Final grade = 0.2 small HAs + 0.2 · group project } more practice
+0.3 · midterm
+0.3 · final } more theory.

Black box →



Beaver
Досёр

Entropy.

Q1. data? No!
Q2. real? Yes!
Q3. box? No!
Q4. serve? yes!
Q5. hedgehog? No!
Q6. Cat? No
Q7. AI? No
Q8. Animal? Yes
Q9. python? No
Q10. U.B? Y
Q11. 4P.M? Y
Q12. dom? N
Q13. Panda? N
Q14. Dail? N
Q15. M? Y

N – number of questions – RV.

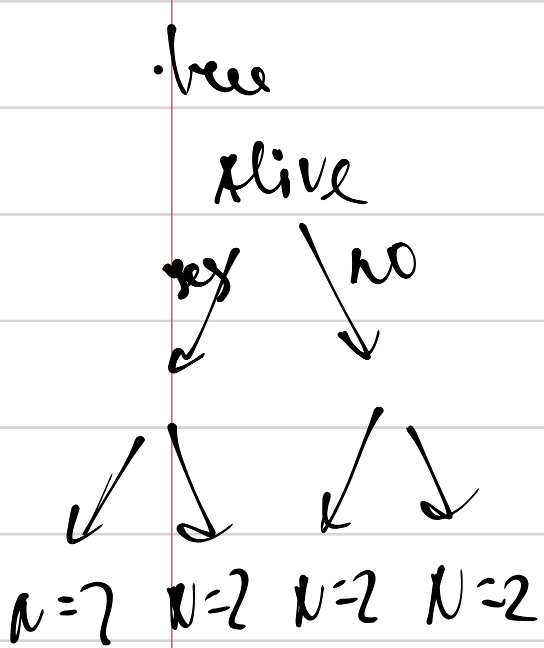
goals:
→ min worst case value of N
→ min E(N) [!]

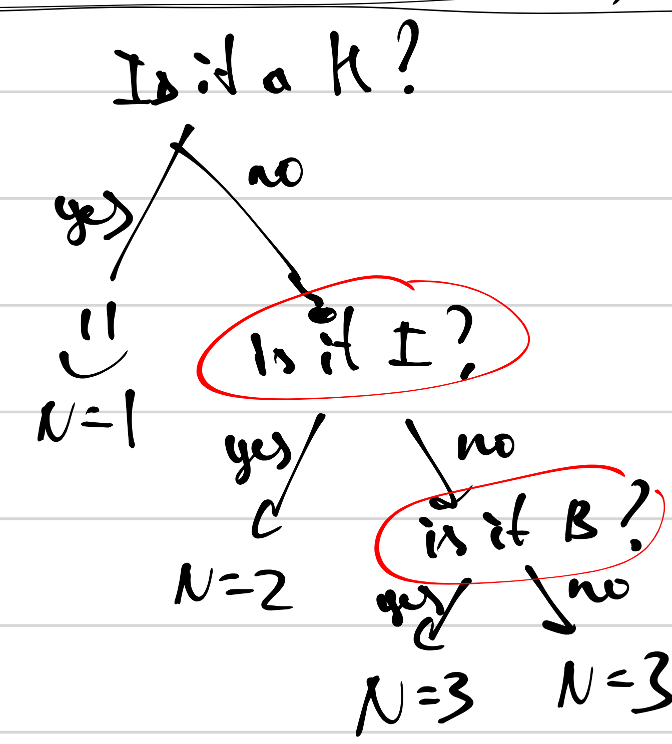
x	Beaver	Orange	Iphone	[illegible]
$P(X=x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$
	$N=3$	$N=3$	$N=2$	$N=1$

$\boxed{\text{strategy}}$ that min $E(N)$?

$p_n = \frac{1}{2^n}$

$n = \log_{\frac{1}{2}} p_n$





worst $N = 2$

$E(N) = 2$

worst $N = 3$

$E(N) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 = 1 + \frac{6}{8} = \boxed{1.75}$

$\boxed{\text{Entropy of } X.}$

$\boxed{\text{Inform}}$ Entropy of R.V X – the expected number of questions to guess the value of X using optimal strategy.

$H(X) = 1.75$

$\boxed{\text{formal}}$

$$H(X) = E(N) = E\left(\log_{\frac{1}{2}} p(X)\right) =$$

where $p(t) = P(X=t)$ [bit]

$$= \boxed{\sum_x P(X=x)\cdot \log_{\frac{1}{2}} P(X=x)}$$

variations: $\log_{\frac{1}{2}} x = -\log_2 x$

$$H(X) = -\sum P(X=x) \cdot \log_2 P(X=x) \quad [bit]$$

$$H(X) = -\sum P(X=x) \cdot \ln P(X=x) \quad [nat]$$

$e = 2.718281828459045....$

nat = NAT

Natural bit

Q. how many bits are there in one nat?

$$H(X) = -\sum P(X=x) \cdot \log_{2^8} P(X=x) \quad [byte]$$

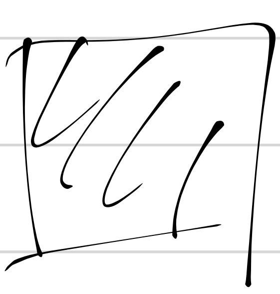
---

y	Y	N
$P(Y=y)$	0.001	0.999

[nat] $H(Y) = -[0.001 \cdot \ln 0.001 + 0.999 \cdot \ln 0.999]$

$= 0.0079.....$

→ idea of many guessing games!



$\underbrace{\text{NNNNNNNNNNNYNNNNNNN...}}_{10^9}$

(informal) $H(X|Q)$ - the av-ge number of questions to guess X if you know the value of Q.

	n	n	n	n	a	a
q / x	I10	I11	I12	I13		
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$H(X|Q) \overset{?}{=} \underbrace{\frac{1}{3}}_{\frac{1}{6}+\frac{1}{6}\ (\text{Alive})} \cdot 1 + \underbrace{\frac{2}{3}}_{\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}\ (\text{Not Alive})} \cdot 2$$

$P(X=x|A) \to$ 0 0 0 0 $\frac{1}{2}$ $\frac{1}{2}$ ($N=1$ $N=1$)

$P(X=x|\text{Not } A) \to$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ 0 0 ($N=2$ $N=2$ $N=2$ $N=2$)

$P(B)\cdot P(A|B) = P(A\cap B)$

formal def.

[nat] $$H(X|Q) = -\sum_q \left\{ P(Q=q) \cdot \sum_x P(X=x|Q=q) \cdot \ln P(X=x|Q=q) \right\}$$

$$= -\sum_{q,x} P(Q=q \cap X=x) \cdot \ln P(X=x|Q=q)$$

$H(X) - H(X|Q) \overset{?}{=}$ meaning? = informational benefit of knowing Q when guessing X

informal $H(X,Y)$ - joint entropy

- expected number of questions to guess X and Y using optimal strategy.

$$H(X,Y) = -\sum_{x,y} P(X=x, Y=y) \cdot \ln P(X=x, Y=y) \quad [\text{nat}]$$

th.

$$H(X,Y) = H(X) + H(Y)$$ if and only if

R.V-s X and Y are independent. [hard]

Th.

$$H(X,Y) \le H(X) + H(Y) \quad \text{(for all } X \text{ and } Y\text{)}$$

Th.

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

[easy]

try to prove this [just plug in the definition]

$$\mathrm{Corr}(X,Y) \neq 0 \Rightarrow X, Y \text{ are dep.}$$

$$\mathrm{Corr}(X,Y) = 0 \rightarrow X, Y \text{ are indep}$$
$$\rightarrow X, Y \text{ are depend}$$

def Joint Information $I(X,Y) = H(X) + H(Y) - H(X,Y)$

$$I(X,Y) = 0 \iff X, Y \text{ are indep}$$

$$I(X,Y) > 0 \iff X, Y \text{ are dep.}$$

real prob.

x	Beaver	Orange	Iphone	[illegible]
$P(X=x)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$
my real prob. (unkn)				
$Q(X=x)$ model	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

optimize str wrt to $Q()$ - prob-ty measure

informal

$CE(P||Q)$ = expected number of questions to guess X if real prob-s are given by P, but you optimize guessing str-gy to Q-probabilities.

cross-entropy from Q to P

th.

$$CE(P||Q) \geq H(P)$$

$$CE(P||P) = H(P)$$

formal

$$CE(P||Q) = -\sum_x P(X=x) \cdot \ln Q(X=x)$$

[nat].