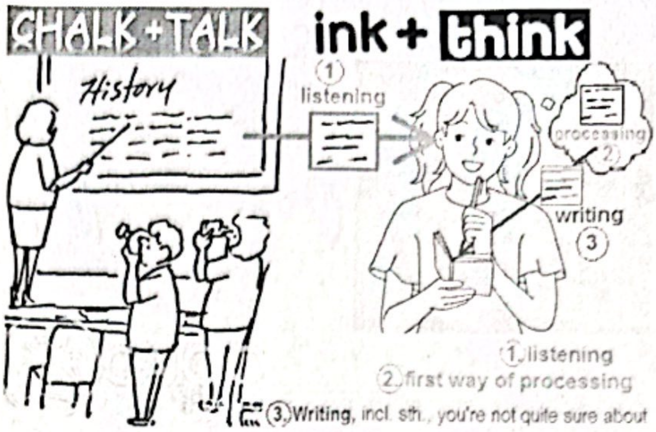


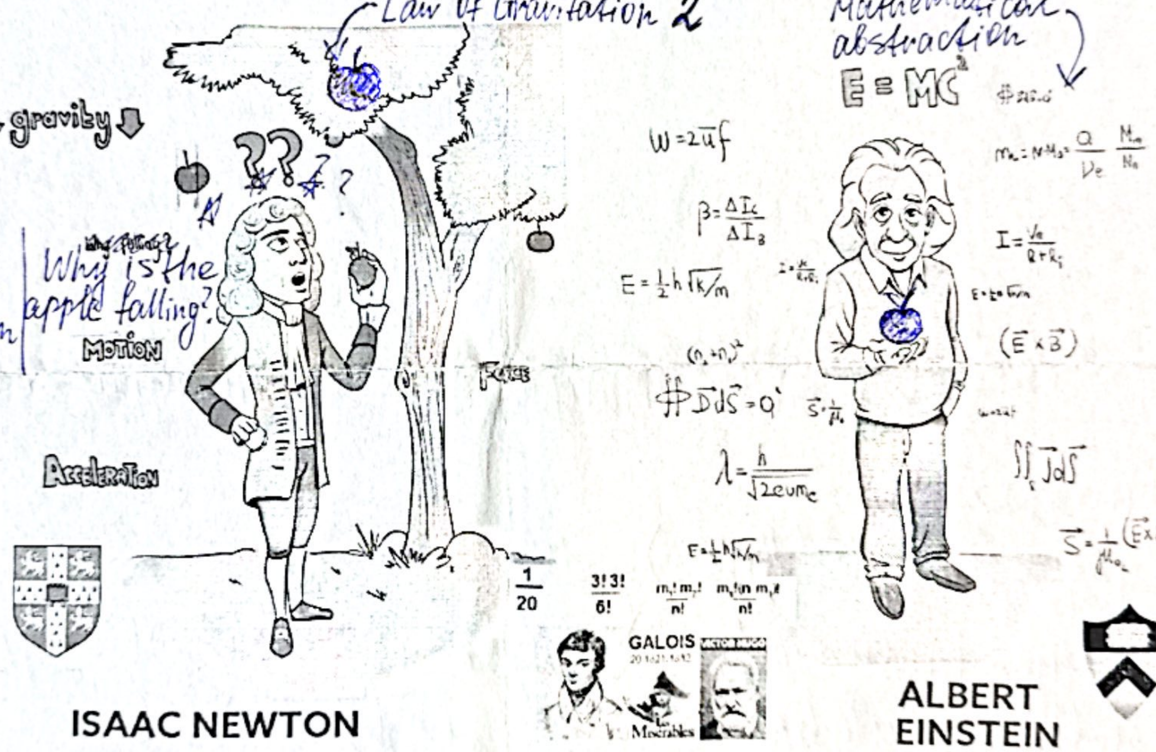
concrete Abstract

School ⇩gravity⇩ MOTION ==formalism==> University $E = MC^2$

CONCRETE AND ABSTRACT THINKING

Concrete thinking ⇩gravity⇩

Physical observation | Why is the apple falling?

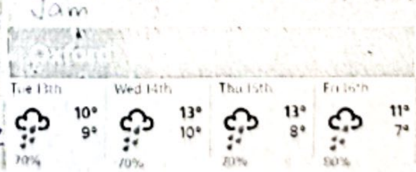


Result: Relativity (Abstract thinking)

Motivation: 80% chance of rain

Let A_j be the event of rain at 9am on day j of this term, $1 \le j \le n$

Suppose the events A_j each have probability p, independently

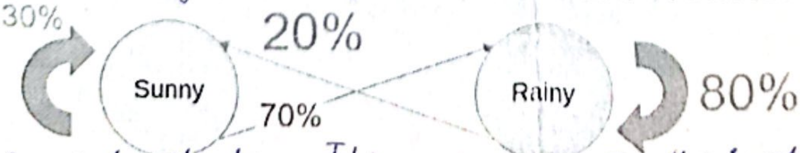


Concrete sees the fact. Abstract sees the probability model.

Markoff Chain Probability Model for Oxford Weather

Formalism: S = Sunny, R = Rainy

$P(S \to S) = 0{,}3$
$P(S \to R) = 0{,}7$
$P(R \to S) = 0{,}2$
$P(R \to R) = 0{,}8$



Sunny → Sunny: 30%
Sunny → Rainy: 70%
Rainy → Sunny: 20%
Rainy → Rainy: 80%
Sum: 100%
Rain chance = 80%

Example: Concrete thinking: It is raining now. I will get wet

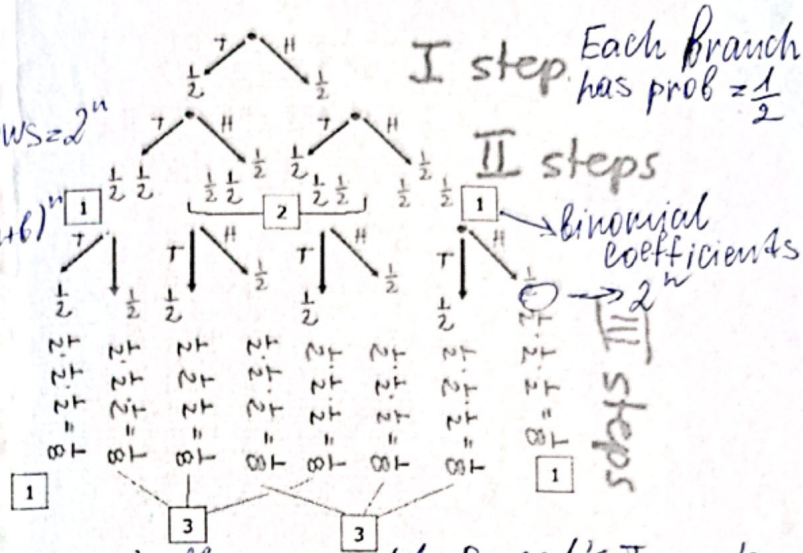
Abstract thinking: Rain today: 80%. Rain tomorrow = 20% if sunny today, 80% if rainy today

Pascal's triangle

1								← 1
1	1							← 2
1	2	1						← 4
1	3	3	1					← 8
1	4	6	4	1				← 16
1	5	10	10	5	1			← 32
1	6	15	20	15	6	1		← 64
1	7	21	35	35	21	7	1	← 128

Sums of rows $= 2^n$

Row n = coeff for $(a+b)^n$



Coefficients match Pascal's Triangle

$$(a+b)^0 = 1$$
$$(a+b)^1 = a+b$$
$$(a+b)^2 = a^2+2ab+b^2$$
$$(a+b)^3 = a^3+3a^2b+3ab^2+b^3$$
$$(a+b)^4 = a^4+4a^3b+6a^2b^2+4ab^3+b^4$$
$$(a+b)^5 = a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5$$

Newton's Binomial

Pascal's Triangle = Newton's Binomial = Prob. Tree

$\left(\frac{1}{2}+\frac{1}{2}\right)^n$

Example:

1. Pascal's Triangle (Row 3): 1, 3, 3, 1
2. Newton's Binomial ($n=3$): $(a+b)^3 = 1a^3+3a^2b+3ab^2+1b^3$
3. Probability Calculation (Were $a=\frac{1}{2}$, $b=\frac{1}{2}$): $1\left(\frac{1}{2}\right)^3 = \frac{1}{8}$

$3\left(\frac{1}{2}\right)^3 = \frac{3}{8}$

$3\left(\frac{1}{2}\right)^3 = \frac{3}{8}$

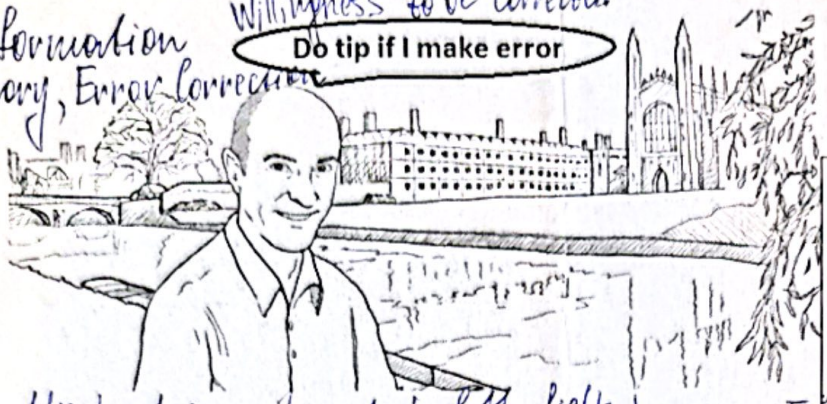
$1\left(\frac{1}{2}\right)^3 = \frac{1}{8}$

Sum = 1

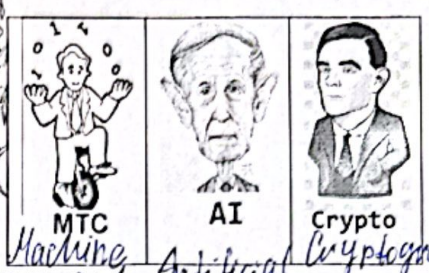
total = 100%

Information theory, Error correction

Willingness to be corrected



Sir **Dr. D. MacKay,**
University of Cambridge
(22 April 1967 – 14 April 2016)



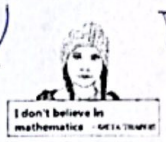
His dual commitment to both fields

Machine Learn. and coding

Artificial Intelligence

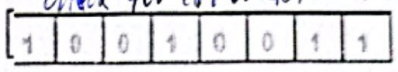
Cryptography

"I believe in clean energy,
but I also believe in mathematics"



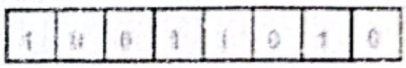
Math is foundation of modern technology.

check for error det

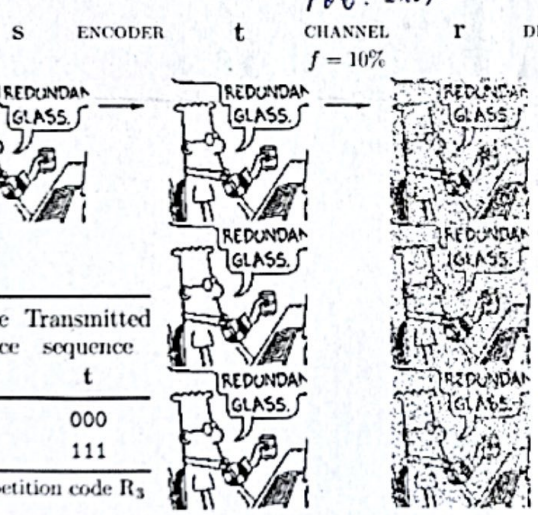


repeating information to survive noise

f=10%



Channel error prob. = 10%



h=noise added in channel
r=receives signal
5 errors found

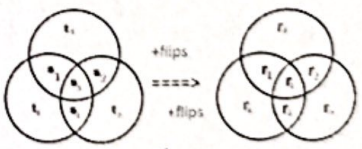
Source sequence s	Transmitted sequence t
0	000
1	111

The repetition code R_3

source message s	0	0	1	0	1	1	0
t	000	000	111	000	111	111	000
n	000	001	000	000	101	000	000
r	000	001	111	000	010	111	000
corrected errors		*			*		
undetected errors							

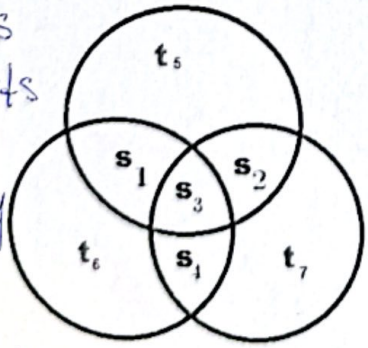
7.4. Hamming code.

$$\frac{4}{\Sigma} \rightarrow \frac{7}{t}$$

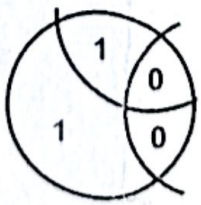


7 bits total
4 data bits
3 parity bits

$\frac{4}{7}$ = efficiency

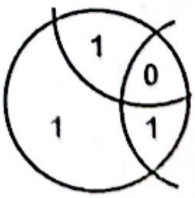


satisfied



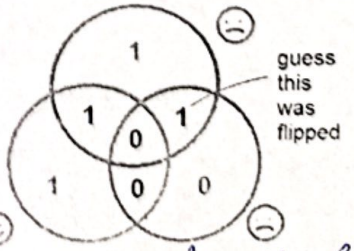
1+1+0+0=0 even

not satisfied



1+1+0+1=1 odd

which circle is wrong?



Formula: $2^m - 1$ bits, correct 1 error

flip the bit in the intersection

+0.1

Strategy: Find the best candidate efficiently

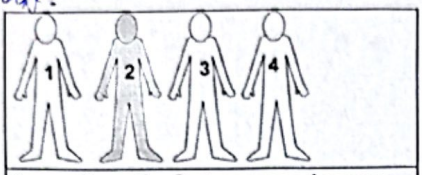
Say NO to the **first**

Say YES to the **second** if it is better than the first

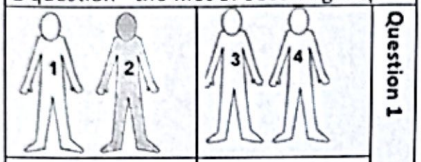
Say NO to the **third** only if it is worse than all the others

Binary Search (Divide into 2 groups each step)

4 people → 2 quest. max.



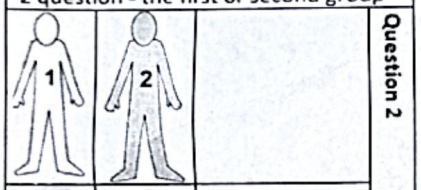
1 question - the first or second group



Question 1

X

2 question - the first or second group



Question 2

X

Average number of questions =

2*0.25+ 2*0.25+ 2*0,25+ 2*0,25 =2

Average number of questions = 1,75

1*0.5+	2*0.25+	3*0,125+	3*0,125
			

Question 1. Is this Zuckerberg?	 50%	1*0.5
Question 2. Is this Sergey Brin?	 25%	2*0.25
Question 3. Is this Stefan from BMW?	 12,5%	3*0,125
So Prince Saud	 12,5%	3*0,125
Average number of questions = 1,75 < 2		

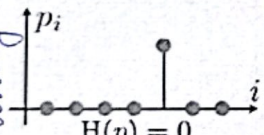
Better than first 2 questions

Average Info.

$$H(X) = \sum_{i=1}^{n} p(x_i) \log_b \frac{1}{p(x_i)}$$

$$\sum_{i=1}^{n} p(i) \log_2 \frac{1}{p(i)}$$

number of bits required on average

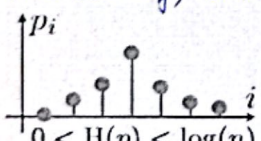


$H(p) = 0$

max certainty, 1 outcome

Quantifying information

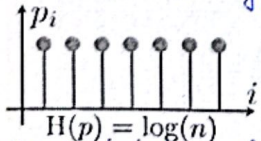
$$I(x_i) = \log_2\left(\frac{1}{p_i}\right)$$



$0 < H(p) < \log(n)$

some uncertainty

number of bits required to encode choice



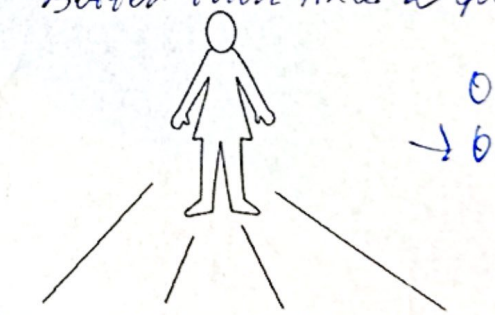
$H(p) = \log(n)$

max uncertainty, uniform distribution

$$\sum_{i=1}^{n} p(x_i)\, I(x_i)$$

0.3 %.o

→ 0,5 %.o



50 billion | 25 billion | 12.5 billion | 12.5 billion

Mark Zukerberg	Sergey Brin	Stefan Quandt	Prince Al Saud
P(1)= 50%	P(2)= 25%	P(3)= 12,5%	P(4)= 12,5%

Sum of probabilities = 100%

5.5 = 25

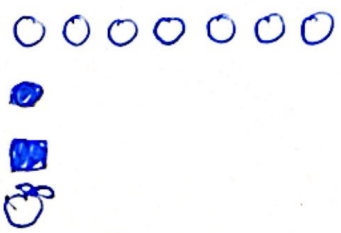
6.6 = 36

7.7 = 49 ↑

5!

1. 0.7

2. 0.1



$p_1 = 0,7$

$p_2 = p_3 = p_4 = 0,1$

$H = -(0,7 \log_2 0,7 + 3 \cdot 0,1 \log_2 0,1) \approx 1,36$ бит

$H = 1 \cdot 0,7 + 2 \cdot 0,1 + 3 \cdot 0,1 + 3 \cdot 0,1 = 0,7 + 0,2 + 0,3 + 0,3 =$

$= 1,5$

Энтропия

Example: Мама мыла раму

Мамамылараму

1. First-order approximation (symbols are independent)
Step-by-step generation:
- Pick 'M' (10% chance) → a
2. Pick 'a' → aa
3. Pick 'M' → aaM
4. Pick 'a' → aaMa
5. Pick a → aaMaa
6. Pick a → aaMaaa
7. Pick a → aaMaaaa
8. Pick т → aaMaaaaт
9. Pick р → aaMaaaaтр
10. Pick а → aaMaaaaтра

Result: aaMaaaaтра
Doesn't look like the orig. phrase at all.

2. Second-order approximation (pares of symbols)

M → only Ma → next a
Ma
a → am(22%), ap(11%) → next
Мам
м → ма(11%) → next т
Мамыт ....

Then we have a result:
Мамыламылараму

M(50%), a(40%), т(10%), л(10%), р(10%)

Ma(22%), am(22%), ма(11%), ам(11%), ла(11%), ар(11%), ра(11%)

Second-order approximation creates meaningful patterns because it remembers the previous letter. First-order is just random noise

First-order approximation (symbols independent but with frequencies of Belarusian txt).

Мама мыла ра

М - 3	30%	1-3 М
а - 4	40%	4-7 а
ы - 1	10%	8 -ы
л - 1	10%	9 -л
р - 1	10%	10 -р

10

лламмар

Мама мыла ра

Ма - 2	22%	1-2	ма
ам - 2	22%	3-4	ам
мы - 1	11%	5	мы
ыл - 1	11%	6	ыл
ла - 1	11%	7	ла
ар - 1	11%	8	ар
ра - 1	11%	9	ра

9

0. 4 6 7 3 1 9 1 6 7 3 5

ам ыл ла ам ма ра ма ыл ла ам мы

мылла рама

More order = better language model

Second-order approximation (digram (2-symbols) structure as in Belarusian)