Edit Distance

⇒ Spelling error correction → word level.

operations → insert
→ delete
→ substitute (character-level)

uniform cost

PART 1

S1 → S2

S2 ** ordered

N → i, insert i,0

2 in

del

S1 INTUITION

1 del
2 in

S2 INSTITUTION

sub s

s s s

2 in

INS

T → S

min

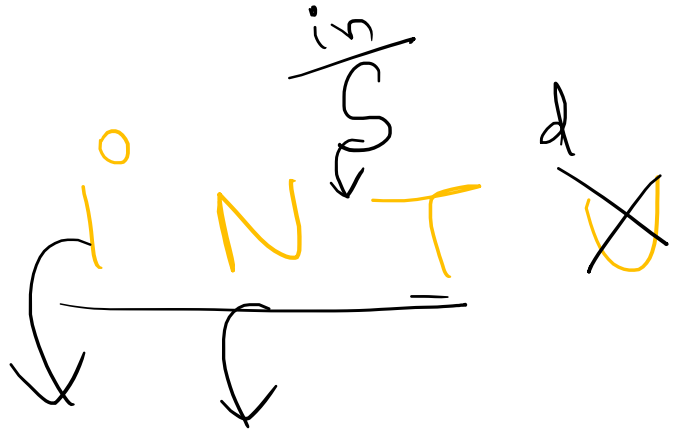
4s / 2 in

5 sub
2 in.

Edit dist = ?

{ Insertion
Del
Subs ?

Cost of each op^n = 1



INTERSECTION

INSTITUTION

3 insertion
1 del
0 substitution

4 edit

General

weighted edit distance =

no. of insertion * cost of insertion

+

no. of deletion * cost of deletion

+

no. of substitution * cost of substitution of

n

(weighted) edit distance = ⇒

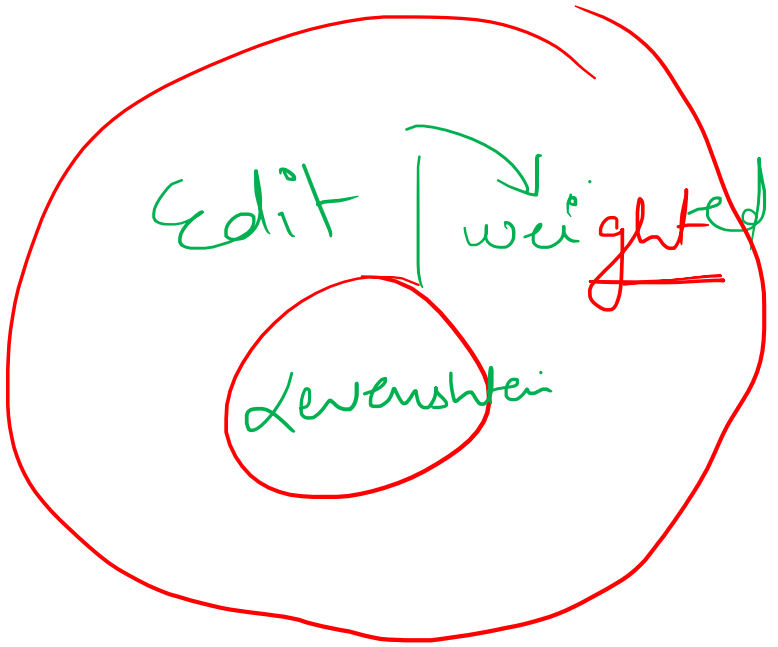
Levenshtein distance ↑

{ cost of deletion = cost of insertion = cost of substitution

cost of each operation = c

total no. of operations = N

$$\sum_{n=1}^{N} n * c$$



Edit distance

insertion : 1

deletion : 1

substitution : 2

⇒ Levenshtein dist

~~opⁿ~~ : cost

insertion : 1

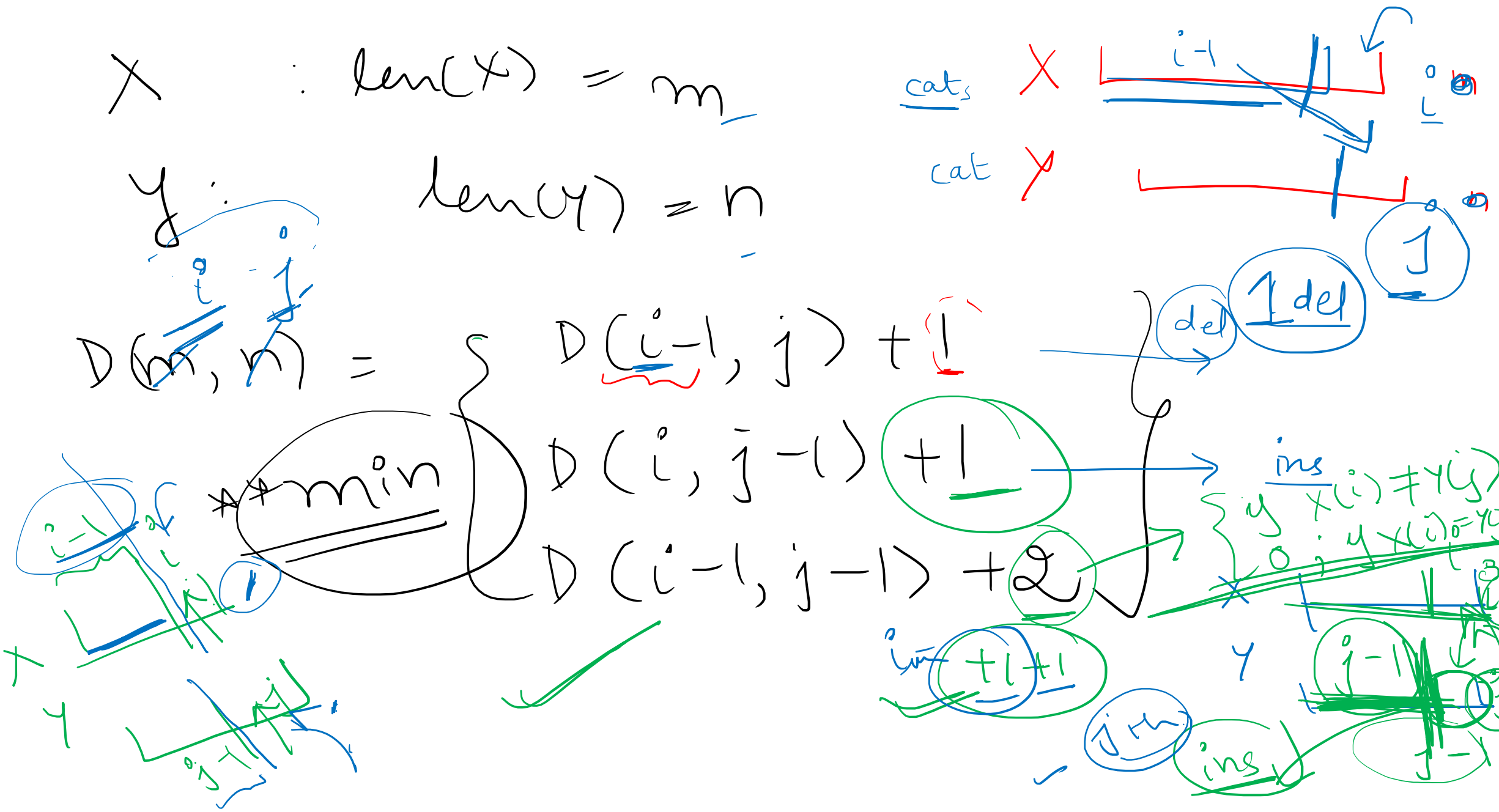
deletion : 1

subs- : 1

X : $len(X) = m$

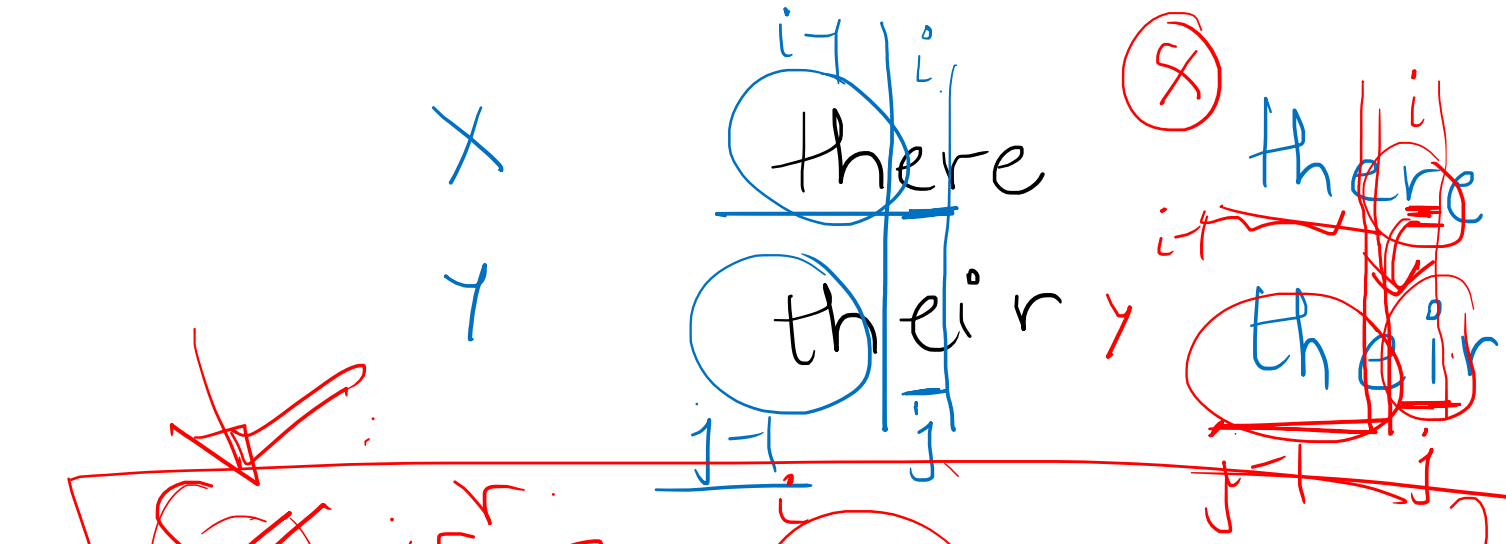
Y : $len(Y) = n$

$$D(m, n) = \min \begin{cases} D(i-1, j) + 1 \\ D(i, j-1) + 1 \\ D(i-1, j-1) + 2 \end{cases}$$



$$D(i,j) = \begin{cases} D(i-1,j)+1 \\ D(i,j-1)+1 \\ D(i-1,j-1)+2 \cdot [X[i] \neq Y[j]] \end{cases}$$

$[X[i] \neq Y[j]]$ — 0; if $X[i] = Y[j]$



c+2 3

c+2 0

Levenshtein

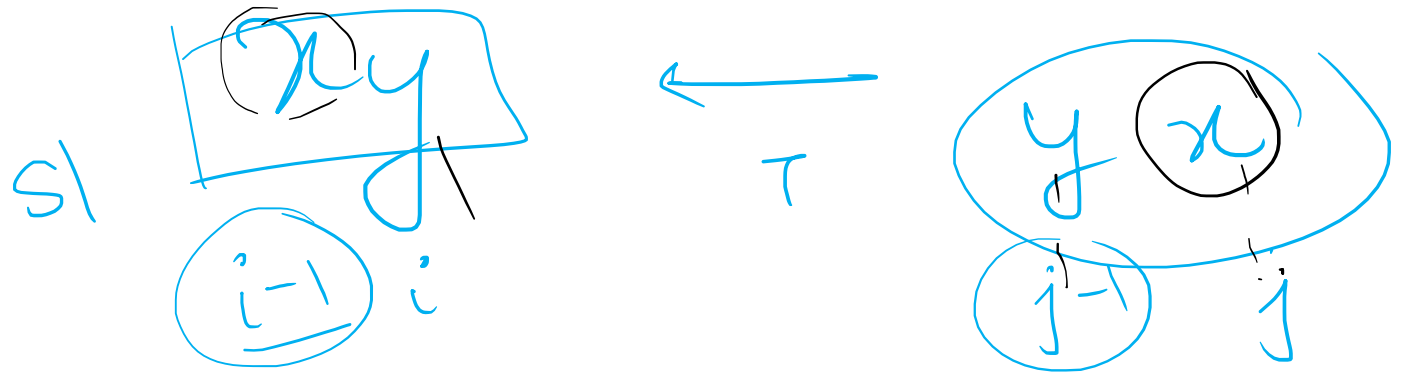
uniform

ins=1
del=1
subs=1

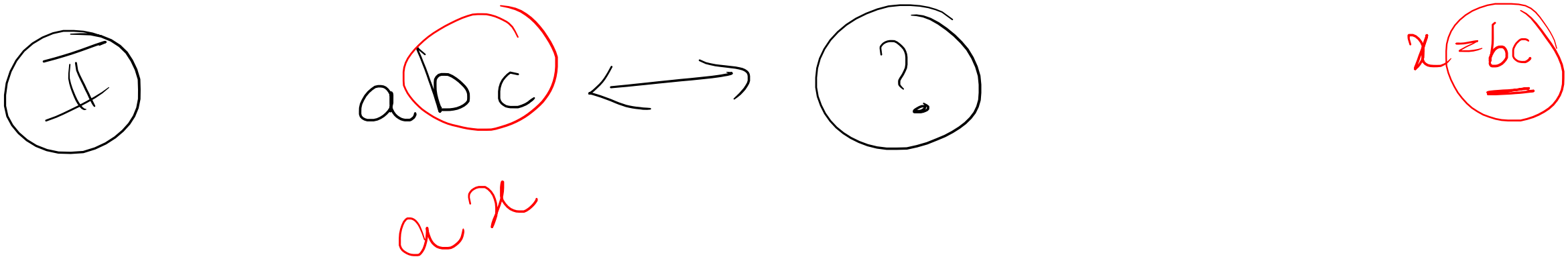
ins = 1 del = 1
subs = 2

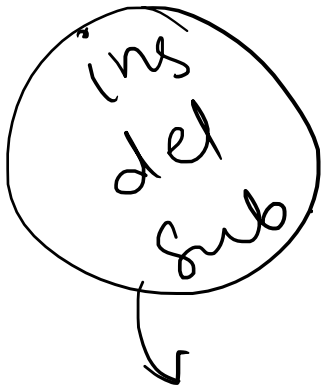
Variation of Edit Distance

(4) → insertion
→ del
→ Substitution
↘ transpose / metathesis.



$$D(i,j) = \begin{cases} s1[i-1] == s2[j] \\ \text{and} \\ s1[i] == s2[j-1] \end{cases}$$

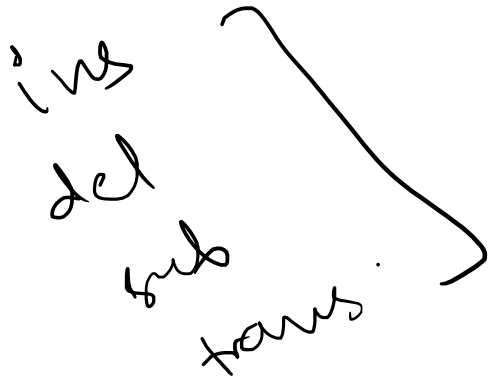

I
abc
?
ax
x=bc

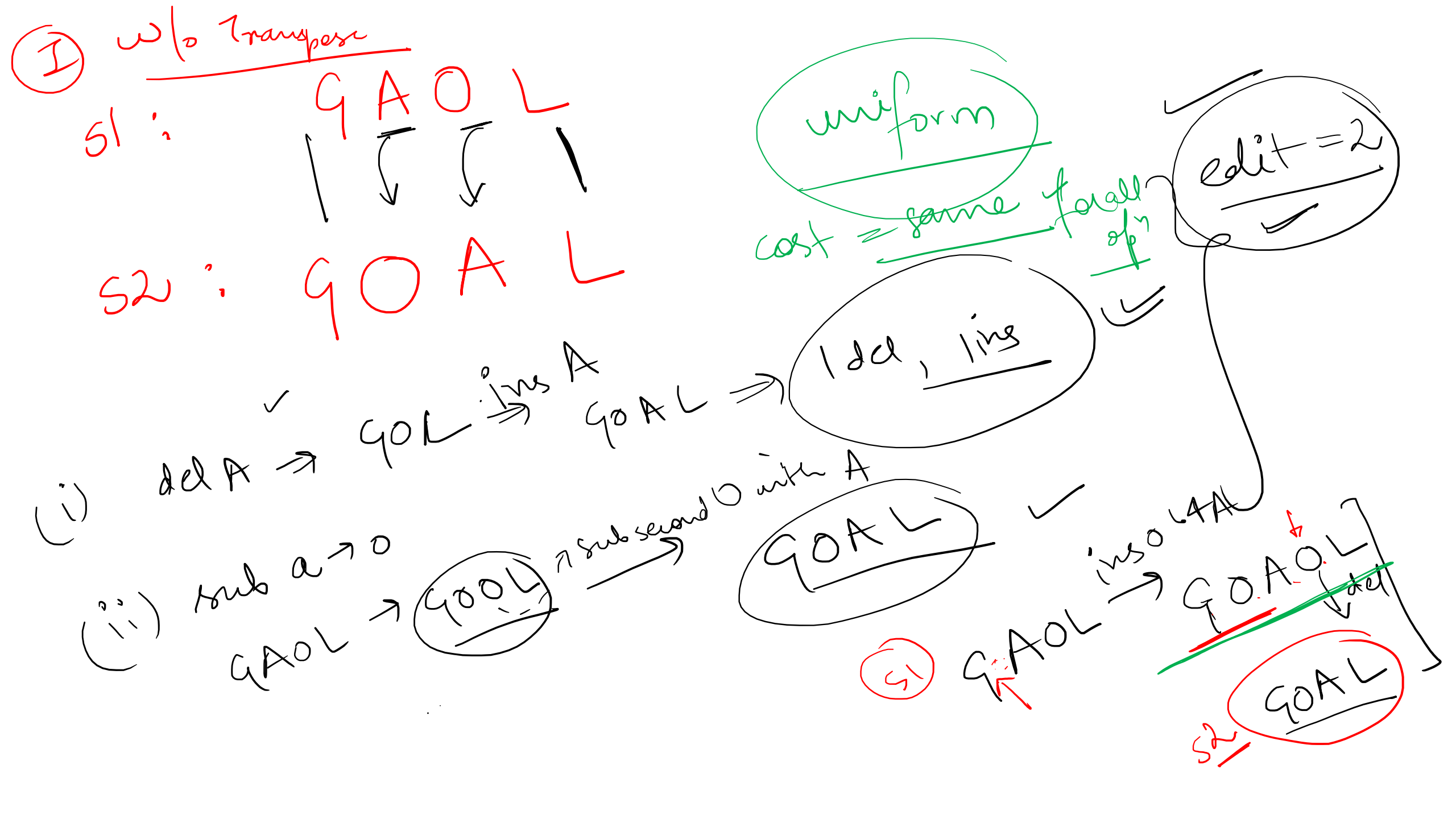


S1: GAOL → S2: GOAL

(i) w/o transposition, edit = ?

(ii) with transposition, edit = ?




① w/o Transpose
S1 : GAOL
S2 : GOAL
uniform
cost = same for all op
edit = 2
(i) del A → GOL ins A → GOAL → Idel, Ins
(ii) sub a → o
GAOL → GOOL → subsequent O with A → GOAL
GAOL → ins O at 4th → GOAOL → del
GOAL
S1
S2

① with transposition

S1: QAOL

S2: QOAL

AO ↔ OA