Dynamic Time Warping

- How similar are EXERCISE and EXIRSAIS ?

E X E R C I S E

E X I R S A I S

d = 5 ?

- ’exercise’ vs. ’zexercises’ ; ’page’ vs. ’ages’

E X E R C I S E

Z E X E R C I S E S

d = ???

P A G E

A G E S

d = ???

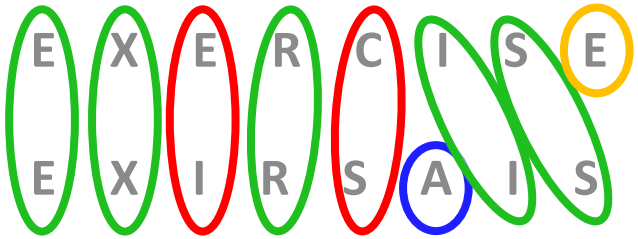
EDIT DISTANCE

Exercise vs. Exirsais

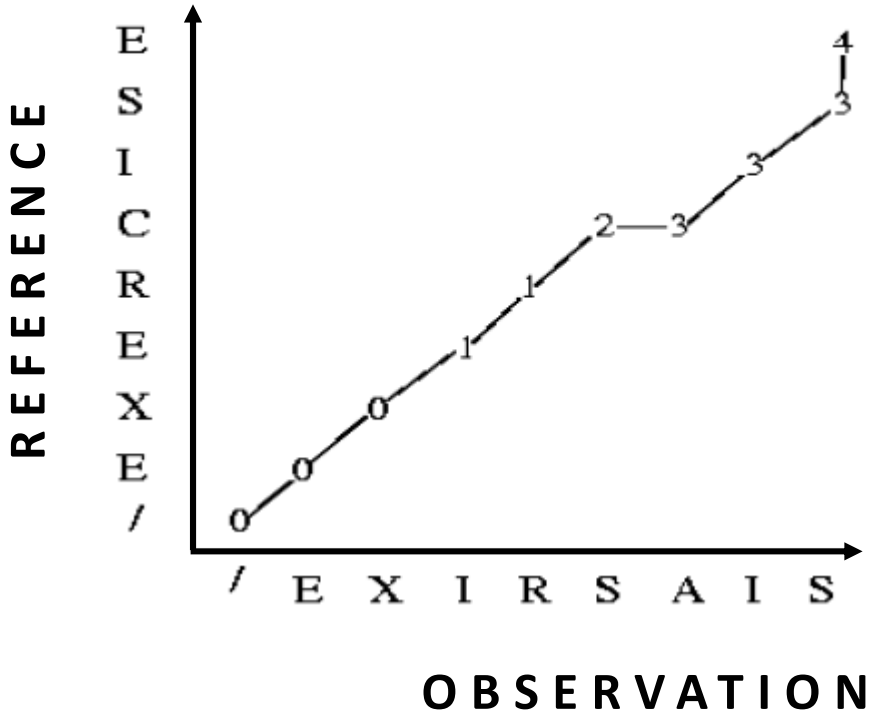
EDIT DISTANCE = A *matching* paradigm for *sequences*

REFERENCE

OBSERVATION



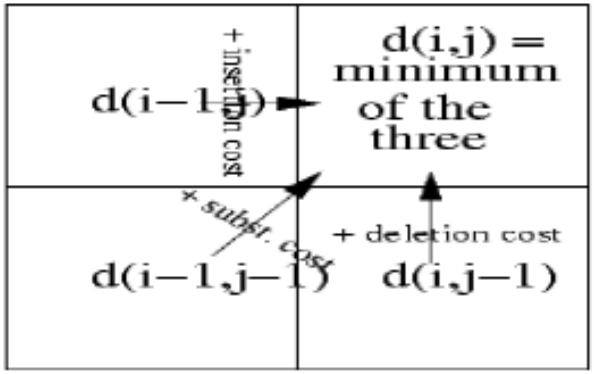
Match
Substitution
Insertion
Deletion



Minimum Edit Distance

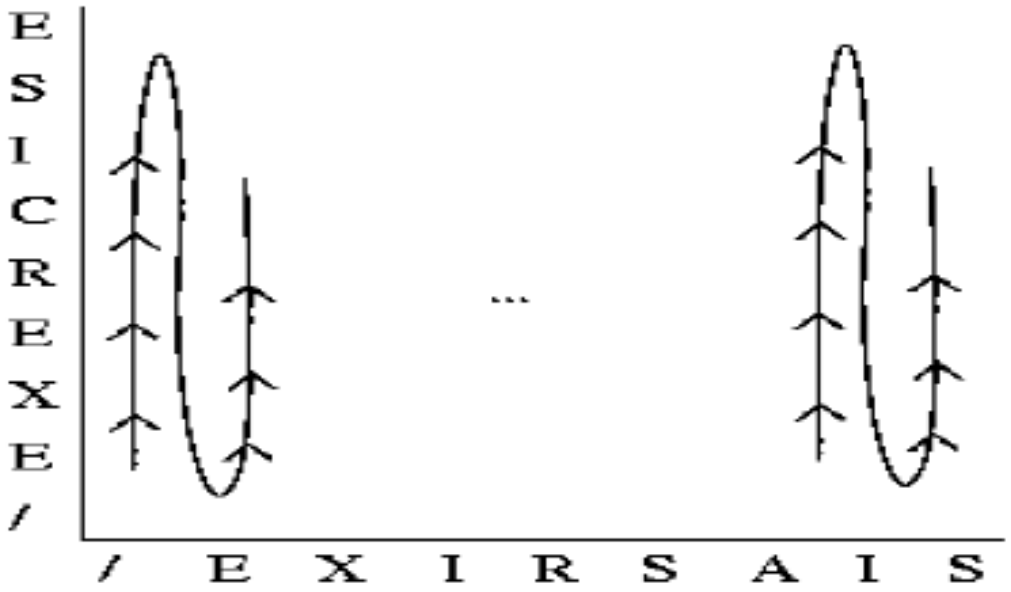
- **MINIMUM EDIT DISTANCE (MED): minimum number of edits that is required to transform REFERENCE to OBSERVATION**
- **The MED can be incrementally computed**
  - **the MED for a partial string edit (reference up to 'j', observation up 'i')**

    = the minimum of {MED for a shorter match + EDIT[$y_j \rightarrow x_i$]}
  - **Incremental Single EDITS**
    - Substitution (cost=0 for matching characters in observation and reference)
    - Deletion (1 character in reference string, no character in observation string)
    - Insertion (no character in reference string, 1 character in observation string)



Trellis Computation

- **Initialization**
  - **d(0,0) = 0**
  - **time needs to move forward for reference and observation**
    - compute column by column
    - compute each column bottom to top

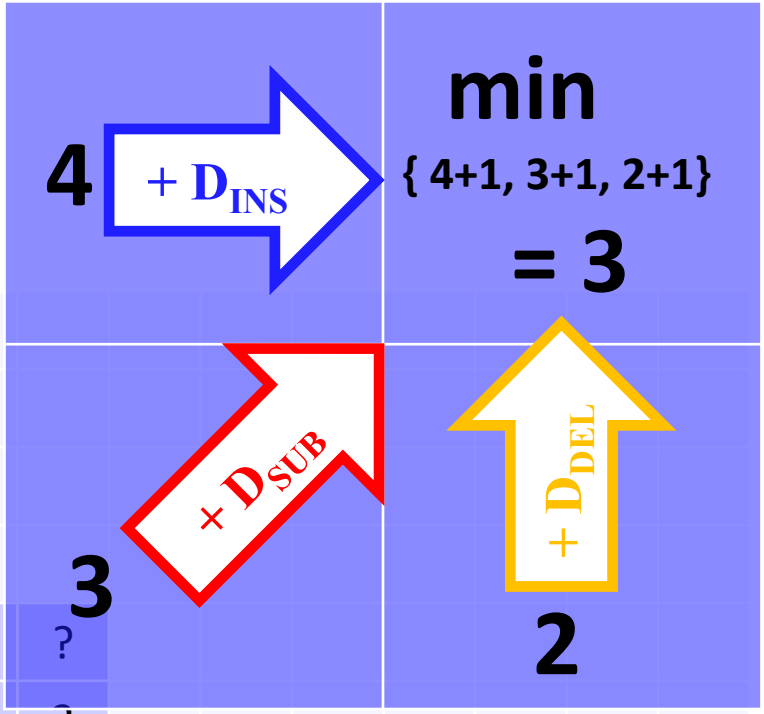


Trellis Computations

From local to accumulated distances

E	1	0	1	1	1	1	1	1	1
S	1	1	1	1	1	0	1	1	0
I	1	1	1	0	1	1	1	0	1
C	1	1	1	1	1	1	1	1	1
R	1	1	1	1	0	1	1	1	1
E	1	0	1	1	1	1	1	1	1
X	1	1	0	1	1	1	1	1	1
E	1	0	1	1	1	1	1	1	1
/	0	1	1	1	1	1	1	1	1
	/	E	X	I	R	S	A	I	S

local distances
(substitution costs)



E	8								
S	7								
I	6								
C	5								
R	4	?							
E	3	2							
X	2	1							
E	1	0							
/	0	1							
	/	E	X	I	R	S	A	I	S

accumulated distances

Trellis Computations
Backpointers, Backtracking and Alignment

local distances

E	1	0	1	1	1	1	1	1	1
S	1	1	1	1	1	0	1	1	0
I	1	1	1	0	1	1	1	0	1
C	1	1	1	1	1	1	1	1	1
R	1	1	1	1	0	1	1	1	1
E	1	0	1	1	1	1	1	1	1
X	1	1	0	1	1	1	1	1	1
E	1	0	1	1	1	1	1	1	1
/	0	1	1	1	1	1	1	1	1
	/	**E**	**X**	**I**	**R**	**S**	**A**	**I**	**S**

accumulated distances

E	8	7	6	5	5	4	4	5	4
S	7	6	5	4	4	3	4	4	3
I	6	5	4	3	3	3	3	3	4
C	5	4	3	3	2	2	3	4	5
R	4	3	2	2	1	2	3	4	5
E	3	2	1	1	2	3	4	4	6
X	2	1	0	1	2	3	4	4	6
E	1	0	1	2	3	4	5	6	7
/	0	1	2	3	4	5	6	7	8
	/	**E**	**X**	**I**	**R**	**S**	**A**	**I**	**S**

Ref	E	X	E	R	C	-	I	S	E
Obs	E	X	I	R	S	A	I	S	-

ALIGNMENT: is found by backtracking, i.e. following backpointers from finishing cell (top+right) to starting cell (bottom+left)

DTW – Questions1

DTW-Q1.1

- **Compute the minimum EDIT distance between**
  - **word (ref) vs. words (obs)** (1)
  - **word (ref) vs. wods (obs)** (2)
  - **words (ref) vs. word (obs)** (1)

DTW-Q1.2

- **Same as DTW-Q1.1, but assume a DELETION cost of 2 for consonants (all other costs are assumed to be '1')**
- **If someone types 'wors', did he/she intend to type 'worse' or 'words' according to your system ?**

DTW – Questions1

DTW-Q1.3

Compute the error rate of a speech recognition system for which you are given the result of a very small test set (Ref = Reference; Hyp = Recognizer output)

S1-Ref: fauchelevent limped along behind the horse in a very contented frame of mind

S1-Hyp: lochleven limped along behind the heard in very contented frame of mind

3/13

S2-Ref: he would have loved to be king in such a non nonsense paradise

S2-Hyp: he had loved the king in a no sense paradigm

8/13

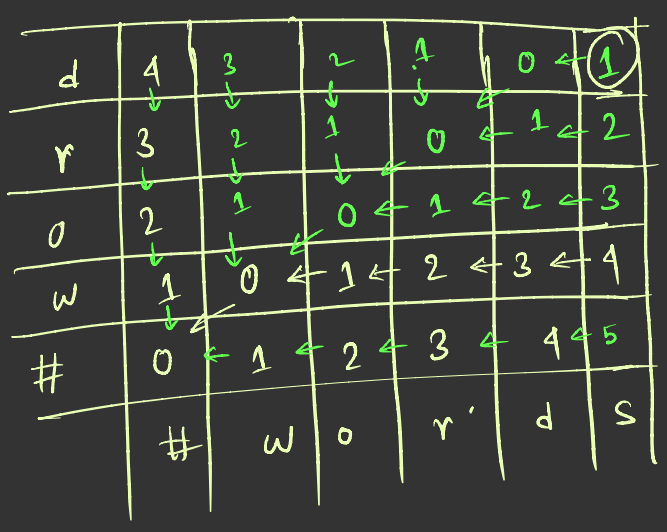
S3-Ref: do you know the names of the seven dwarfs in Disney's Snow White movie ?

S3-Hyp: do you know the names of the seven warfs in the sneaze now white movie ?

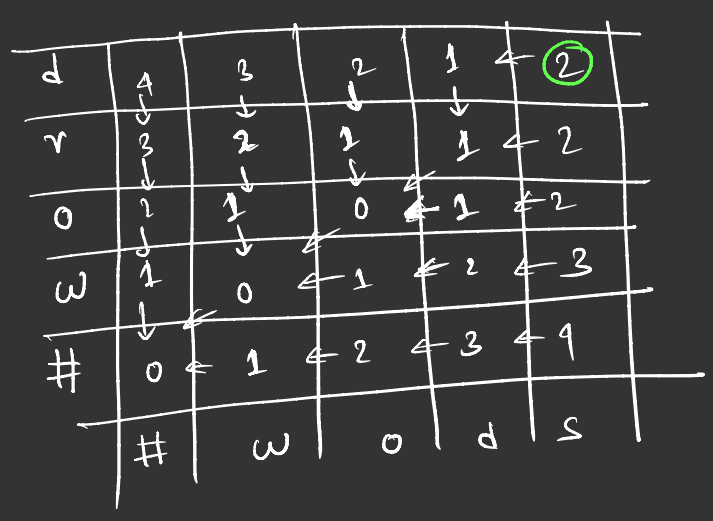
5/15

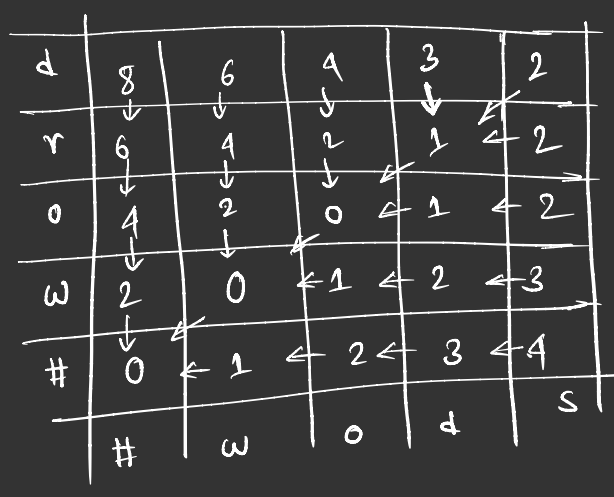
ERR: 16/41

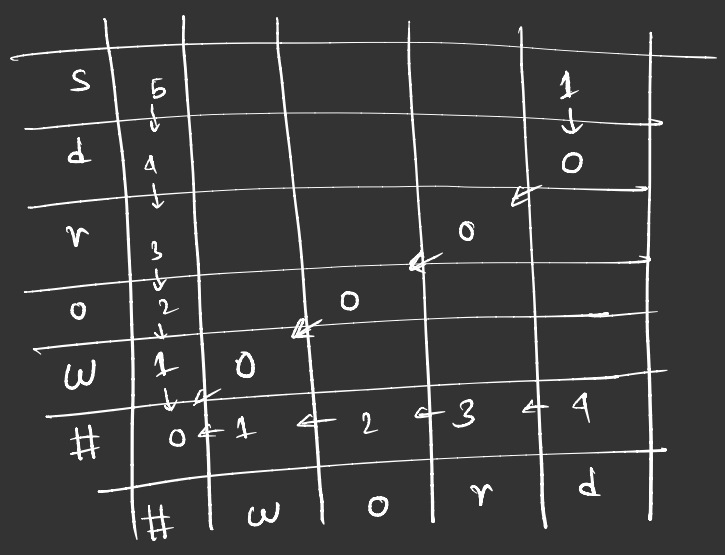
1.1

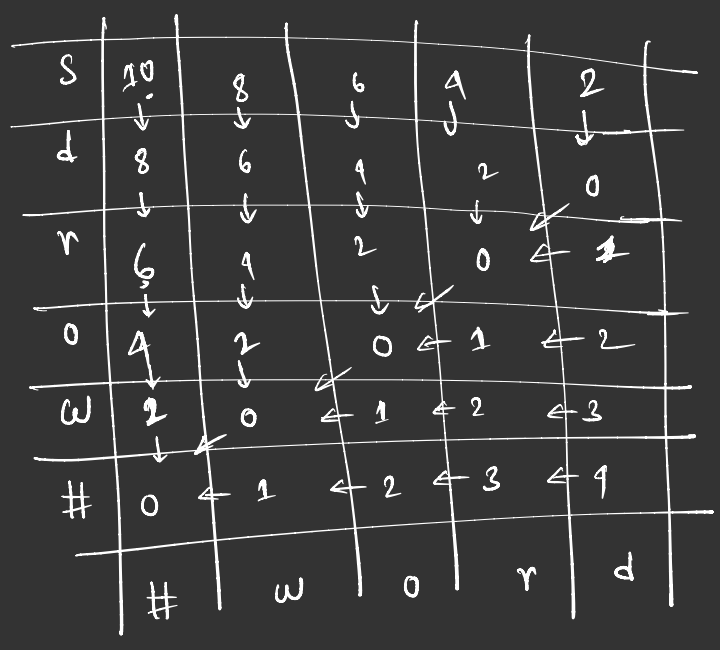


word
wods









1.3 wors → morse