

1)

1)

A) Levenshtein Distance / Algorithm:

$L_{ij}(x, y)$: Edit distance between the first i element of x and the first j elements of y

$$L_{ij} = \begin{cases} i & j=0 \\ j & i=0 \\ \min \{ L_{i-1,j} + 1, L_{i,j-1} + 1, L_{i-1,j-1} + p(x^{(i)}, y^{(j)}) \} & \text{otherwise} \end{cases}$$

$$\text{where } p(x, y) = \begin{cases} 0 & \text{if } x=y \\ 1 & \text{otherwise} \end{cases}$$

Pseudocode:

for $s = 1, \dots, i$ $L_{s0} = s$ for $t = 1, \dots, j$ $L_{s0} = t$ for $s = 1, \dots, i$ for $t = 1, \dots, j$ $L_{st} = \min \{ L_{s-1,t} + 1,$ $L_{s,t-1} + 1,$ $L_{s-1,t-1} + p(x^{(s)}, y^{(t)}) \}$

Edit distance [verification]

HORSE
↓ ↓ ↓ ↓ ↓
⑤ SORSE

∴ edit distance is 1

S = Substitution

Using the 3 insertion, deletion & substitution we can obtain the edit distance.

Here we have H → S the rest remains the same.

~~Verification~~ Verification matrix

	S	O	R	S	E	
O	1	2	3	4	5	
H	1	2	3	4	5	6
O	2	3	2		4	5
R	3	4	3			
S	4	5	4	3	2	3
E	5	6	5	4	3	2

	S	O	R	S	E	
O	1	2	3	4	5	
H	1	1	2	3	4	5
O	2	2	1	2	3	4
R	3	3	2	1	2	3
S	4	3	3	2	1	2
E	5	4	4	3	2	1

edit distance is ①

Application of edit distance:

- × It's application is that it can be used if it is used in spell check (Auto correct)
- × It is used to find the distance between two strings
- × It is also used in the dictionary to display all the words in that ^{are} near proximity.

* If the s_1 & s_2 don't match.

(1) S & H don't match hence.

$$\min(1, 0, 1) + 1 = 1 \quad (0+1)$$

(2) S & O don't match.

$$\min(1, 1, 2) + 1 = 2.$$

(3) S & R don't match.

$$\min(2, 2, 3) + 1 = 3.$$

(4) S & S match. hence the diagonal element
are common

(5) S & i don't match.

$$\min(3, 4, 0) + 1 = 4.$$

there fore after using the algorithm we know that the we need
1 string to be filled with another to obtain the final output.

2)

there fore after using the algorithm we know that the we need
1 string to the rules with another to obtain the favorable output .

2) Feature Selection:

Feature selection is the process where we automatically
manually select those features / fields which contribute
most to your prediction variables or output analysis
if we select irrelevant features in the field / ~~file~~ in the
data it reduces the accuracy of the model & make's
it irrelevant features .

Importance of FS:

- * It enables the machine learning algorithm to learn faster.
- * It reduces the complexity of a model & makes it easier to interpret.
- * It improves the accuracy of a model if the right subset is chosen.
- * It reduces overfitting.

Relevant features - those that ^{we} need to perform well.

Irrelevant features - those that are simply unnecessary.

Redundant features - those that become irrelevant in the presence of others.

* 3 main categories of feature selection

in the presence of others

4 3 main categories of feature selection.

- Filter method
- Wrapper method
- Embedded method

Conclusion on feature selection:-

Practical benefits:

- To improve accuracy
- Reduce computation
- Reduce space.
- Reduce cost of feature measurement
- Improved data / model understanding

