

Learning objectives

- What is autocorrect?
- Building the model
- Minimum edit distance
- Minimum edit distance algorithm

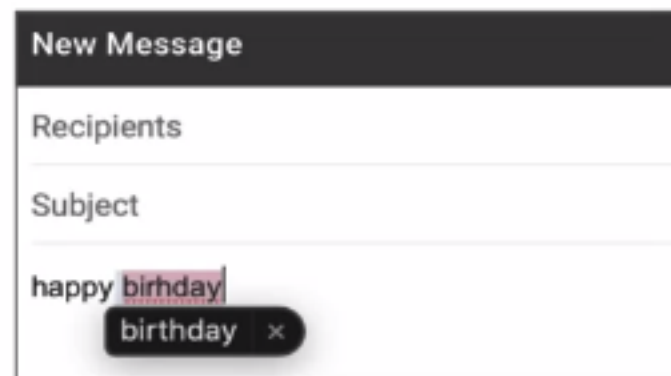
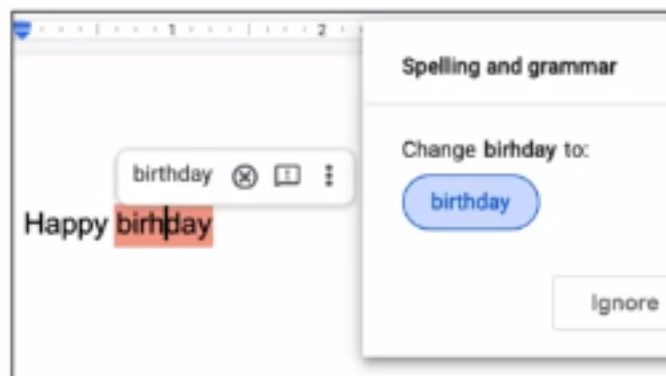
deah → dear ✓

yeah
dear
dean
... etc

	#	s	t	a	y
#	0	1	2	3	4
p	1	2	3	4	5
l	2	3	4	5	6
a	3	4	5	4	5
y	4	5	6	5	4

What is autocorrect?

- Phones
- Tablets
- Computers



What is autocorrect?

- Example:

Happy birthday deah friend!



What is autocorrect ?

- Example:

Happy birthday  dear friend! 

What is autocorrect?

- Example:

Happy birthday deer friend!  ??

How it works

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

How it works

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

deah

How it works

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

deah

_eah

d_ar

de_r

... etc

How it works

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

deah
yeah
dear
dean
... etc

How it works

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

deah

yeah

dear

dean

... etc


How it works

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

deah → dear ✓
yeah
|dear|
dean
... etc

Building the model


1. Identify a misspelled word

deah ?? 

Building the model

1. Identify a misspelled word

```
if word not in vocab:  
    misspelled = True
```

deah ?? 

Building the model

1. Identify a misspelled word

```
if word not in vocab:  
    misspelled = True
```

deah



deer



Building the model

1. Identify a misspelled word

```
if word not in vocab:  
    misspelled = True
```

deah



Happy birthday deer !



Building the model

2. Find strings n edit distance away

Building the model

2. Find strings n edit distance away
 - Edit: an operation performed on a string to change it

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter)

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter)

'to': 'top', 'two' ...

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter) 'to': 'top', 'two' ...
- Delete (remove a letter)

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter)
- Delete (remove a letter)

'to': 'top', 'two' ...

'hat': 'ha', 'at', 'ht'

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter) 'to': 'top', 'two' ...
- Delete (remove a letter) 'hat': 'ha', 'at', 'ht'
- Switch (swap 2 adjacent letters)

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter) 'to': 'top', 'two' ...
- Delete (remove a letter) 'hat': 'ha', 'at', 'ht'
- Switch (swap 2 adjacent letters) 'eta': 'eat', 'tea'

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it

- Insert (add a letter)

'to': 'top', 'two' ...

- Delete (remove a letter)

'hat': 'ha', 'at', 'ht'

- Switch (swap 2 adjacent letters)

'eta': 'eat', 'tea'

'ate' ❌

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter) 'to': 'top', 'two' ...
- Delete (remove a letter) 'hat': 'ha', 'at', 'ht'
- Switch (swap 2 adjacent letters) 'eta': 'eat', 'tea'
- Replace (change 1 letter to another)

Building the model

2. Find strings n edit distance away

- Edit: an operation performed on a string to change it
- Insert (add a letter) 'to': 'top', 'two' ...
- Delete (remove a letter) 'hat': 'ha', 'at', 'ht'
- Switch (swap 2 adjacent letters) 'eta': 'eat', 'tea'
- Replace (change 1 letter to another) 'jaw': 'jar', 'paw', ...

Building the model

2. Find strings n edit distance away
 - Given a string find all possible strings that are n edit distance away using
 - Input
 - Delete
 - Switch
 - Replace

Building the model

2. Find strings n edit distance away
 - Given a string find all possible strings that are n edit distance away using
 - Input
 - Delete
 - Switch
 - Replace

deah
_eah
d_ar
de_r
... etc

Building the model

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

Building the model

3. Filter candidates

deah
_eah
d_ar
de_r
... *etc*

Building the model

3. Filter candidates

deah
_eah
d_ar
de_r
... etc

→

deah
yeah
dear
dean
... etc

Building the model

1. Identify a misspelled word
2. Find strings n edit distance away
3. Filter candidates
4. Calculate word probabilities

Building the model

4. Calculate word probabilities

Building the model

4. Calculate word probabilities

Example: "I am happy because I am learning"

Word	Count
I	2
am	2
happy	1
because	1
learning	1

Total: 7

Building the model

4. Calculate word probabilities

Example: "I am happy because I am learning"

Word	Count
I	2
am	2
happy	1
because	1
learning	1

Total: 7

Building the model

4. Calculate word probabilities

Example: "I am happy because I am learning"

Word	Count
I	2
am	2
happy	1
because	1
learning	1

Total: 7

Building the model

4. Calculate word probabilities

Example: "I am happy because I am learning"

Word	Count
I	2
am	2
happy	1
because	1
learning	1

Total: 7

Building the model

4. Calculate word probabilities

Example: "I am happy because I am learning"

$$P(w) = \frac{C(w)}{V}$$

$P(w)$ Probability of a word

$C(w)$ Number of times the word appears

V Total size of the corpus

Word	Count
I	2
am	2
happy	1
because	1
learning	1

Total: 7

Building the model

4. Calculate word probabilities

Example: "I am happy because I am learning"

$$P(w) = \frac{C(w)}{V}$$

$$P(\text{am}) = \frac{C(\text{am})}{V} = \frac{2}{7}$$

$P(w)$ Probability of a word

$C(w)$ Number of times the word appears

V Total size of the corpus

Word	Count
I	2
am	2
happy	1
because	1
learning	1

Total: 7

Building the model

4. Calculate word probabilities

deah
yeah
dear
dean
... etc

Building the model

4. Calculate word probabilities

deah → dear ✓
yeah
dear
dean
... etc

Summary

1. Identify a misspelled word
2. Find strings n edit distance away

Insert
Delete
Switch
Replace

3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear 

yeah

dear

dean

... etc

Summary

1. Identify a misspelled word
2. Find strings n edit distance away

Insert
Delete
Switch
Replace

3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear ✓
yeah
dear
dean
... etc

Summary

1. Identify a misspelled word
2. Find strings n edit distance away
 - Insert
 - Delete
 - Switch
 - Replace
3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear ✓

yeah

dear

dean

... etc

Summary

1. Identify a misspelled word
2. Find strings n edit distance away

Insert
Delete
Switch
Replace

3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear ✓
_eah
d_ar
de_r
... etc

Summary

1. Identify a misspelled word
2. Find strings n edit distance away

Insert
Delete
Switch
Replace

3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear ✓
yeah
dear
dean
... etc

Summary

1. Identify a misspelled word
2. Find strings n edit distance away

Insert
Delete
Switch
Replace

3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear 

yeah

dear

dean

... etc

Summary

1. Identify a misspelled word
2. Find strings n edit distance away

Insert
Delete
Switch
Replace

3. Filter candidates
4. Calculate word probabilities

$$P(w) = \frac{C(w)}{V}$$

deah → dear ✓

yeah
dear
dean
... etc

Minimum edit distance

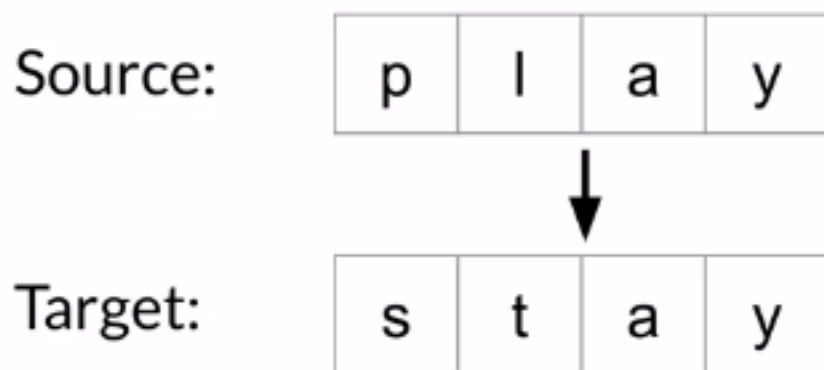
- How to evaluate similarity between 2 strings?
- Minimum number of edits needed to transform 1 string into the other
- Spelling correction, document similarity, machine translation, DNA sequencing, and more

Minimum edit distance

- Edits:
- Insert (add a letter) 'to': 'top', 'two' ...
- Delete (remove a letter) 'hat': 'ha', 'at', 'ht'
- Replace (change 1 letter to another) 'jaw': 'jar', 'paw', ...

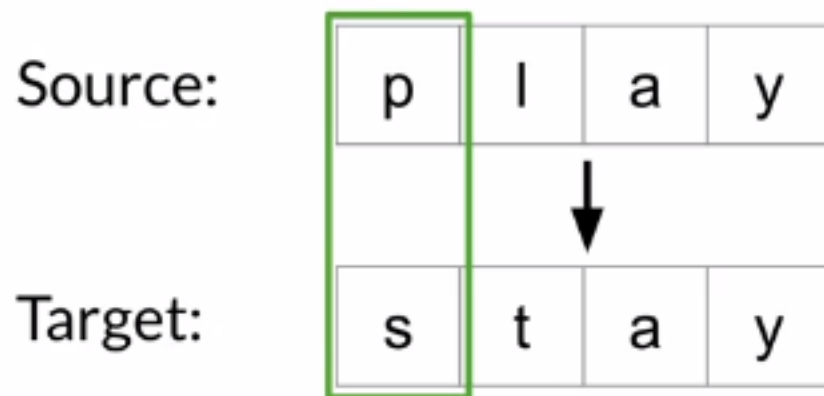
Minimum edit distance

- Example:



Minimum edit distance

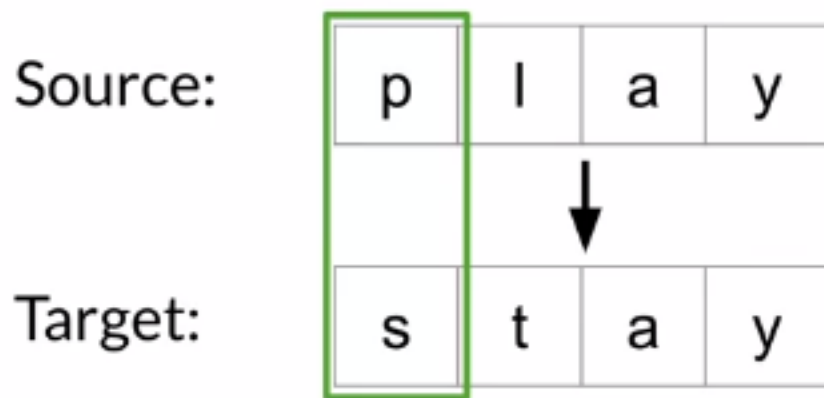
- Example:



$p \rightarrow s$: replace

Minimum edit distance

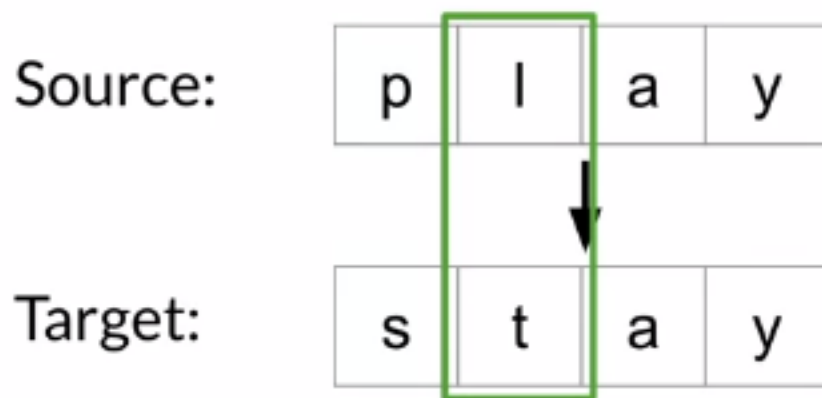
- Example:



$p \rightarrow s$: replace

Minimum edit distance

- Example:

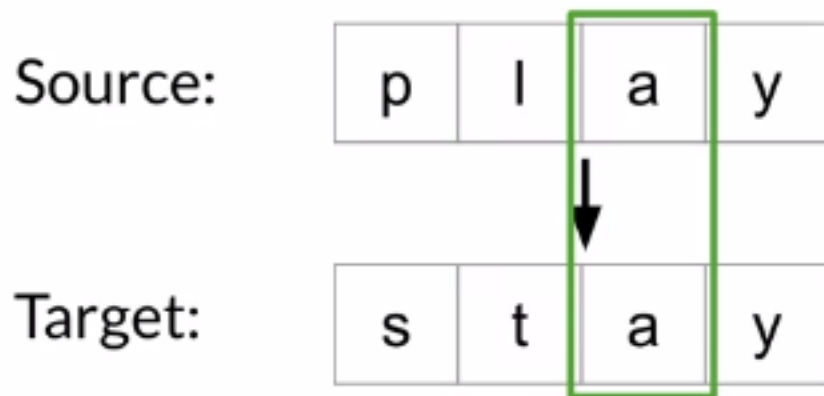


$p \rightarrow s$: replace

$l \rightarrow t$: replace

Minimum edit distance

- Example:

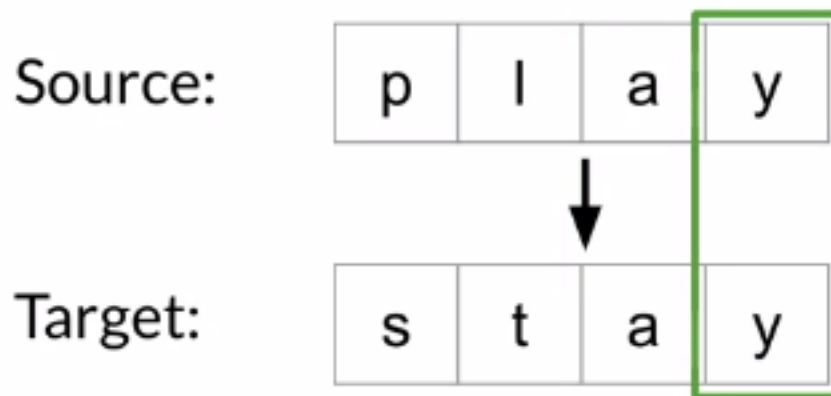


$p \rightarrow s$: replace

$l \rightarrow t$: replace

Minimum edit distance

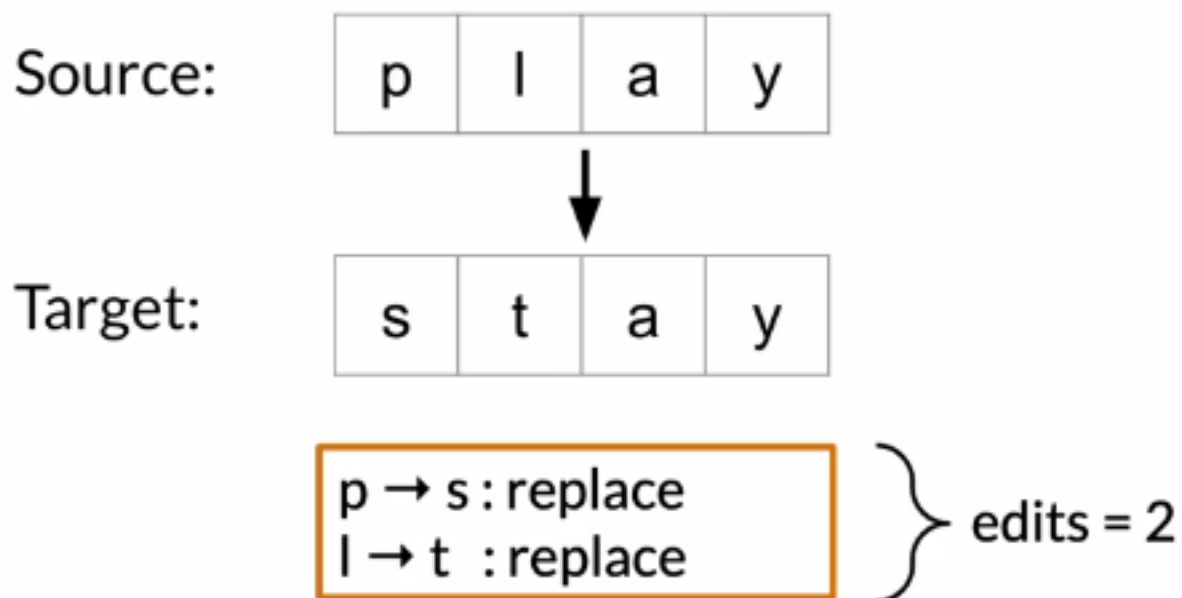
- Example:



p → s : replace
l → t : replace

Minimum edit distance

- Example:



Minimum edit distance

- Example:

Source:

p	l	a	y
---	---	---	---



Target:

s	t	a	y
---	---	---	---

Edit cost:

Insert 1

Delete 1

Replace 2

p → s : replace
l → t : replace

} edits = 2

Minimum edit distance

- Example:

Source:

p	l	a	y
---	---	---	---



Target:

s	t	a	y
---	---	---	---

Edit cost:

Insert 1

Delete 1

Replace 2

$$\text{edit distance} = 2 * 2 = 4$$

p → s : replace
l → t : replace

} edits = 2

Minimum edit distance

- Example:

c	o	n	v	o	l	u	t	i	o	n	a	l	n	e	u	r	a	l	n	e	t	w	o	r	k
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Minimum edit distance

- Example:

c	o	n	v	o	l	u	t	i	o	n	a	l	n	e	u	r	a	l	n	e	t	w	o	r	k
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

CCAAGGGGTGACTCTAGTTTAATACTGAGATCAAATTATATGGGTGAT



!!

Minimum edit distance

Source: play → Target: stay

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

	0	1	2	3	4
	#	s	t	a	y
0	#				
1	p				
2	l				
3	a				
4	y				

Minimum edit distance

Source: play → Target: stay

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

$D[]$

	0	1	2	3	4
	#	s	t	a	y
0	#				
1	p				
2	l				
3	a				
4	y				

Minimum edit distance

Source: play → Target: stay

$D[]$

$D[2,3] = pl \rightarrow sta$

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

$D[]$

$D[2,3] = pl \rightarrow sta$

$D[2,3] = \text{source}[:2] \rightarrow \text{target}[:3]$

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

$D[]$

$D[2,3] = \text{pl} \rightarrow \text{sta}$

$D[2,3] = \text{source}[:2] \rightarrow \text{target}[:3]$

$D[i,j] = \text{source}[:i] \rightarrow \text{target}[:j]$

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

$D[]$

$D[i,j] = \text{source}[:i] \rightarrow \text{target}[:j]$

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play \rightarrow Target: stay

$D[]$

$D[i,j] = \text{source}[:i] \rightarrow \text{target}[:j]$

$D[m,n] = \text{source} \rightarrow \text{target}$

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

$D[]$

$D[i,j] = \text{source}[:i] \rightarrow \text{target}[:j]$

$D[m,n] = \text{source} \rightarrow \text{target}$

	0	1	2	3	4
	#	s	t	a	y
0	#				
1	p				
2	l				
3	a				
4	y				

The diagram illustrates the alignment between the source string 'play' and the target string 'stay' in a dynamic programming table. The source string is represented by the first column (indices 0-4) and the target string by the first row (indices 0-4). The alignment path is highlighted with orange arrows: a horizontal arrow from (0,0) to (0,2), a vertical arrow from (0,0) to (2,0), and a diagonal arrow from (0,0) to (2,2). This indicates that the minimum edit distance involves deleting 'p' and 'l' and inserting 's' and 't'.

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

		0	1	2	3	4
		#	s	t	a	y
0	#					
1	p					
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

→

	0	1	2	3	4
	#				
0	#				
1					
2					
3					
4					

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → #

		0	1	2	3	4
		#				
0	#	0				
1	p					
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → #
delete

		0	1	2	3	4
		#				
0	#	0				
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

→ s

		0	1	2	3	4
		#	s			
0	#	0				
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

→ s
insert

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

$p \rightarrow s$

insert + delete: $p \rightarrow ps \rightarrow s$

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

$p \rightarrow s$

insert + delete: $p \rightarrow ps \rightarrow s$

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

$p \rightarrow s$

insert + delete: $p \rightarrow ps \rightarrow s$: 2

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

$p \rightarrow s$

insert + delete: $p \rightarrow ps \rightarrow s$: 2

delete + insert: $p \rightarrow \# \rightarrow s$

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

$p \rightarrow s$

insert + delete: $p \rightarrow ps \rightarrow s$: 2

delete + insert: $p \rightarrow \# \rightarrow s$

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

insert + delete: p → ps → s: 2

delete + insert: p → # → s: 2

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

insert + delete: p → ps → s: 2

delete + insert: p → # → s: 2

replace: p → s

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

insert + delete: p → ps → s: 2

delete + insert: p → # → s: 2

replace: p → s: 2

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1				
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

insert + delete: p → ps → s: 2

delete + insert: p → # → s: 2

replace: p → s: 2

		0	1	2	3	4
		#	s			
0	#	0	1			
1	p	1	2			
2						
3						
4						

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1			
1	p	1	2			
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

play → #

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1			
1	p	1	2			
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

play → #

$$D[i, j] = D[i-1, j] + del_cost$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1			
1	p	1	2			
2	l					
3	a					
4	y					

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

play → #

$$D[i, j] = D[i-1, j] + del_cost$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1			
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

play → #

$$D[i, j] = D[i-1, j] + del_cost$$

$$D[4, 0] = \text{play} \rightarrow \# \\ = \text{source}[:4] \rightarrow \text{target}[0]$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1			
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

→ play

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1			
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

→ play

$$D[i, j] = D[i, j-1] + ins_cost$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1			
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

→ play

$$D[i, j] = D[i, j-1] + ins_cost$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$D[i, j] =$

$$\min \begin{cases} D[i-1, j] + del_cost \\ D[i, j-1] + ins_cost \\ D[i-1, j-1] + \begin{cases} rep_cost; & \text{if } src[i] \neq tar[j] \\ 0; & \text{if } src[i] = tar[j] \end{cases} \end{cases}$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$D[i, j] =$

$$\min \begin{cases} D[i-1, j] + \text{del_cost} \\ D[i, j-1] + \text{ins_cost} \\ D[i-1, j-1] + \begin{cases} \text{rep_cost}; & \text{if } \text{src}[i] \neq \text{tar}[j] \\ 0; & \text{if } \text{src}[i] = \text{tar}[j] \end{cases} \end{cases}$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$D[i, j] =$

$$\min \begin{cases} D[i-1, j] + \text{del_cost} \\ D[i, j-1] + \text{ins_cost} \\ D[i-1, j-1] + \begin{cases} \text{rep_cost}; & \text{if } \text{src}[i] \neq \text{tar}[j] \\ 0; & \text{if } \text{src}[i] = \text{tar}[j] \end{cases} \end{cases}$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$D[i, j] =$

$$\min \begin{cases} D[i-1, j] + \text{del_cost} \\ D[i, j-1] + \text{ins_cost} \\ D[i-1, j-1] + \begin{cases} \text{rep_cost}; & \text{if } \text{src}[i] \neq \text{tar}[j] \\ 0; & \text{if } \text{src}[i] = \text{tar}[j] \end{cases} \end{cases}$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$D[i, j] =$

$$\min \begin{cases} D[i-1, j] + \text{del_cost} \\ D[i, j-1] + \text{ins_cost} \\ D[i-1, j-1] + \begin{cases} \text{rep_cost}; & \text{if } \text{src}[i] \neq \text{tar}[j] \\ 0; & \text{if } \text{src}[i] = \text{tar}[j] \end{cases} \end{cases}$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$D[i, j] =$

$$\min \begin{cases} D[i-1, j] + \text{del_cost} \\ D[i, j-1] + \text{ins_cost} \\ D[i-1, j-1] + \begin{cases} \text{rep_cost}; & \text{if } \text{src}[i] \neq \text{tar}[j] \\ 0; & \text{if } \text{src}[i] = \text{tar}[j] \end{cases} \end{cases}$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$$D[i-1, j] + 1 = 2$$

$$D[i, j-1] + 1 = 2$$

$$D[i-1, j-1] + 2 = 2$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$$D[i-1, j] + 1 = 2$$

$$D[i, j-1] + 1 = 2$$

$$D[i-1, j-1] + 2 = 2$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

p → s

$$D[i-1, j] + 1 = 2$$

$$D[i, j-1] + 1 = 2$$

$$D[i-1, j-1] + 2 = 2$$

min = 2

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2			
2	l	2				
3	a	3				
4	y	4				

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

play → stay

$$D[m, n] = 4$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2	3	4	5
2	l	2	3	4	5	6
3	a	3	4	5	4	5
4	y	4	5	6	5	4

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

play → stay

$D[m, n] = 4$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2	3	4	5
2	l	2	3	4	5	6
3	a	3	4	5	4	5
4	y	4	5	6	5	4

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

play → stay

$$D[m, n] = 4$$

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2	3	4	5
2	l	2	3	4	5	6
3	a	3	4	5	4	5
4	y	4	5	6	5	4

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

- Levenshtein distance

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2	3	4	5
2	l	2	3	4	5	6
3	a	3	4	5	4	5
4	y	4	5	6	5	4

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

- Levenshtein distance
- Backtrace

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2	3	4	5
2	l	2	3	4	5	6
3	a	3	4	5	4	5
4	y	4	5	6	5	4

Minimum edit distance

Source: play → Target: stay

Cost: insert: 1, delete: 1, replace: 2

- Levenshtein distance
- Backtrace
- Dynamic programming

		0	1	2	3	4
		#	s	t	a	y
0	#	0	1	2	3	4
1	p	1	2	3	4	5
2	l	2	3	4	5	6
3	a	3	4	5	4	5
4	y	4	5	6	5	4

Summary - learning objectives

- What is autocorrect ?
- Building the model
- Minimum edit distance
- Minimum edit distance algorithm

deah → dear ✓

yeah

dear

dean

... etc

	#	s	t	a	y
#	0	1	2	3	4
p	1	2	3	4	5
l	2	3	4	5	6
a	3	4	5	4	5
y	4	5	6	5	4