

# Expanding window features: part 1

---

Window features

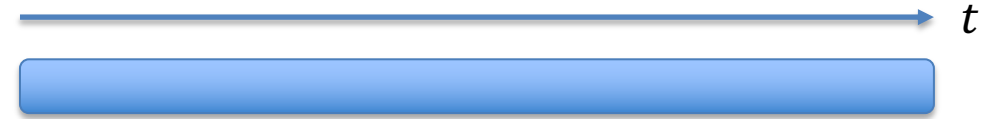
# Rolling vs expanding features

Rolling window



- Uses only some of the previous values at any step.
- Useful to capture recent behaviors of the time series.
- Example: the average sales over the past one month.

Expanding window



- Uses all previous values at any step.
- Useful when you need access to the entire time series.
- Examples:
  - aggregating a variable which has a cumulative effect.
  - target encoding.

# Expanding window features

Date	Sales
2020-02-12	35
2020-02-13	30
2020-02-14	23
2020-02-15	30
2020-02-16	34
2020-02-17	?

- Apply a window to the time series.

# Expanding window features

Date	Sales
2020-02-12	35
2020-02-13	30
2020-02-14	23
2020-02-15	30
2020-02-16	34
2020-02-17	?

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.

# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	35		
2020-02-13	30		
2020-02-14	23		
2020-02-15	30		
2020-02-16	34		
2020-02-17	?		

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.

# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	<b>35</b>		
2020-02-13	30	<b>35.0</b>	<b>35</b>
2020-02-14	23		
2020-02-15	30		
2020-02-16	34		
2020-02-17	?		

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.

# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	<b>35</b>		
2020-02-13	30	<b>35.0</b>	<b>35</b>
2020-02-14	23		
2020-02-15	30		
2020-02-16	34		
2020-02-17	?		

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.
- **Expand** window and iterate across the time series.

# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	<b>35</b>		
2020-02-13	<b>30</b>	35.0	35
2020-02-14	23	<b>32.5</b>	<b>30</b>
2020-02-15	30		
2020-02-16	34		
2020-02-17	?		

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.
- **Expand** window and iterate across the time series.



# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	<b>35</b>		
2020-02-13	<b>30</b>	35.0	35
2020-02-14	<b>23</b>	32.5	30
2020-02-15	30	<b>29.3</b>	<b>23</b>
2020-02-16	34		
2020-02-17	?		

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.
- **Expand** window and iterate across the time series.

# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	<b>35</b>		
2020-02-13	<b>30</b>	35.0	35
2020-02-14	<b>23</b>	32.5	30
2020-02-15	<b>30</b>	29.3	23
2020-02-16	34	<b>29.5</b>	<b>23</b>
2020-02-17	?		

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.
- **Expand** window and iterate across the time series.

# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	<b>35</b>		
2020-02-13	<b>30</b>	35.0	35
2020-02-14	<b>23</b>	32.5	30
2020-02-15	<b>30</b>	29.3	23
2020-02-16	<b>34</b>	29.5	23
2020-02-17	<b>?</b>	<b>30.4</b>	<b>23</b>

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.
- **Expand** window and iterate across the time series.

# Expanding window features

Date	Sales	Sales mean	Sales min
2020-02-12	35	NaN	NaN
2020-02-13	30	35.0	35
2020-02-14	23	32.5	30
2020-02-15	30	29.3	23
2020-02-16	34	29.5	23
2020-02-17	?	30.4	23

- Apply a window to the time series.
- Compute statistics from data inside the window (e.g., mean, min).
- To avoid data leakage, assign the statistics to timestamp after window.
- **Expand** window and iterate across the time series.