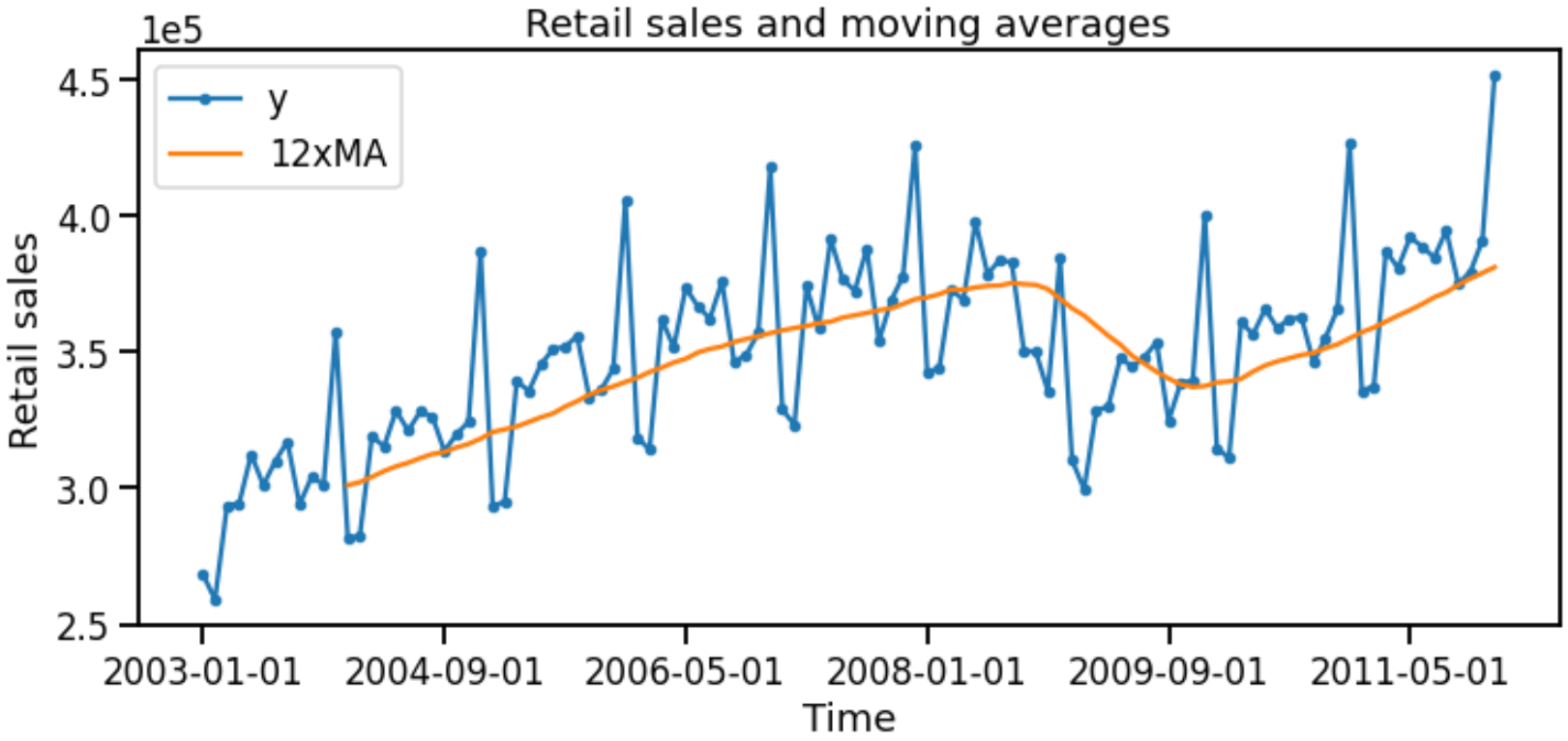
Weighted window functions: part 1

Window features

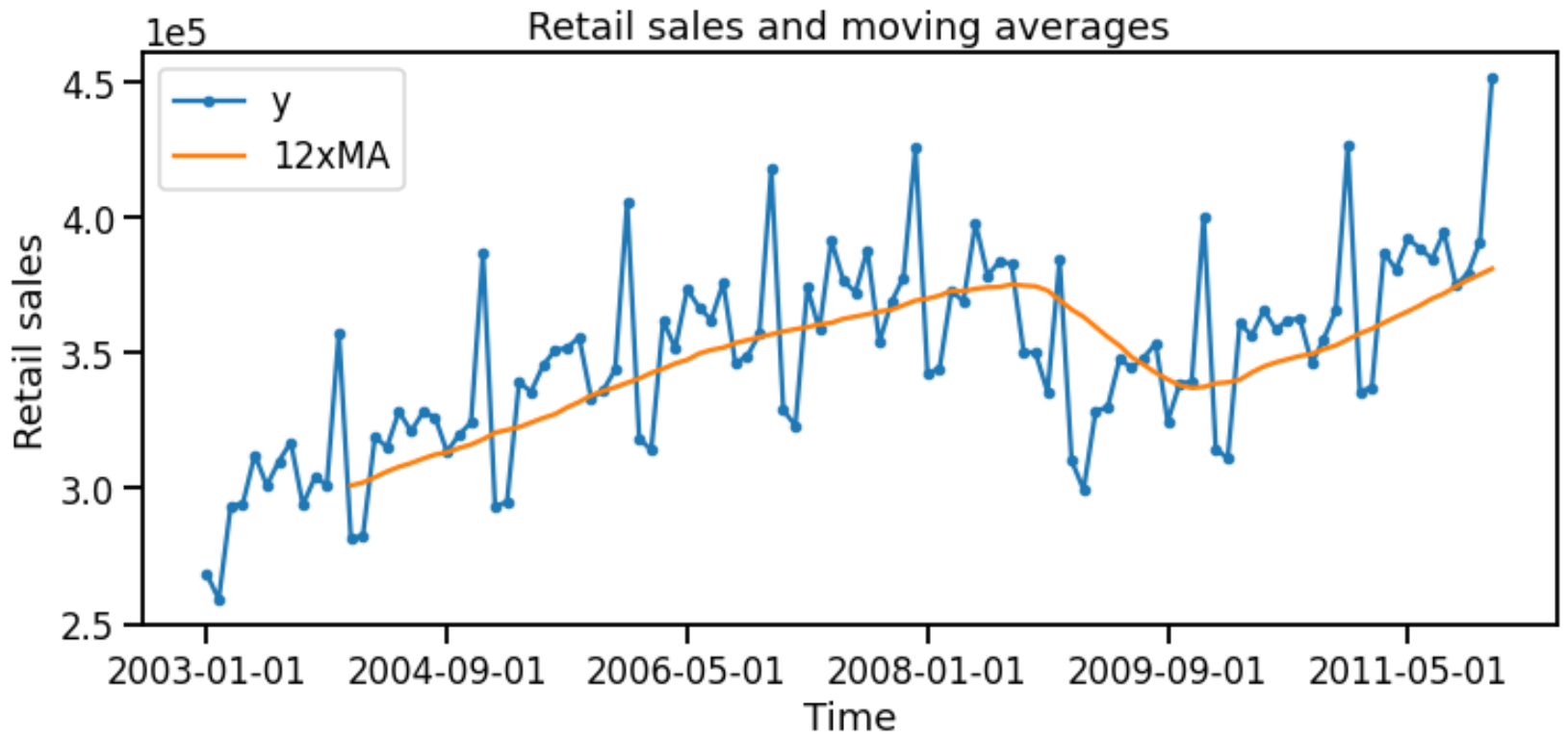
Motivation for weighted window functions

- What if we want to be more sensitive to recent observations, e.g., to quickly pick up changes in trend?



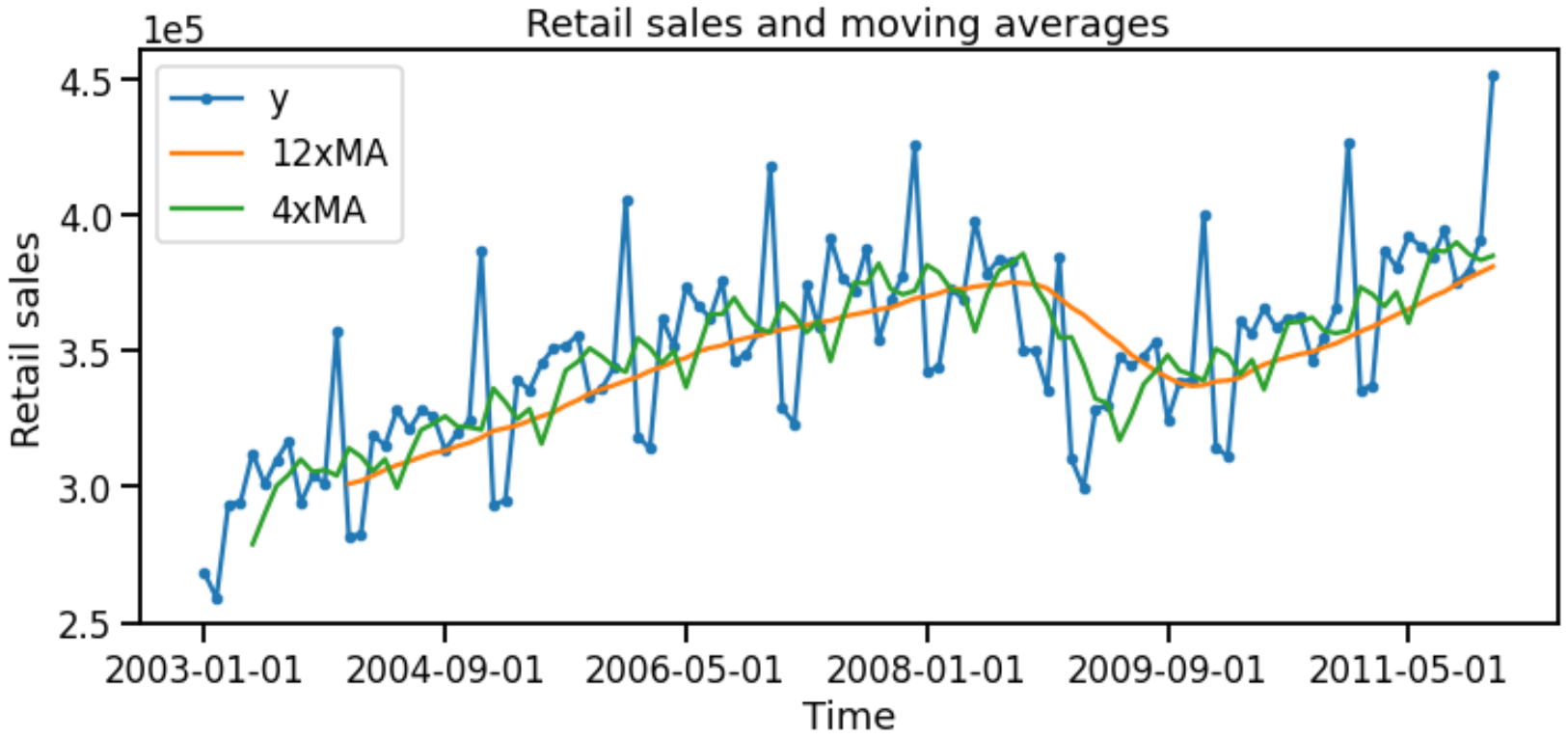
Motivation for weighted window functions

- We could use a shorter window but this would increase the variance of the new window feature.



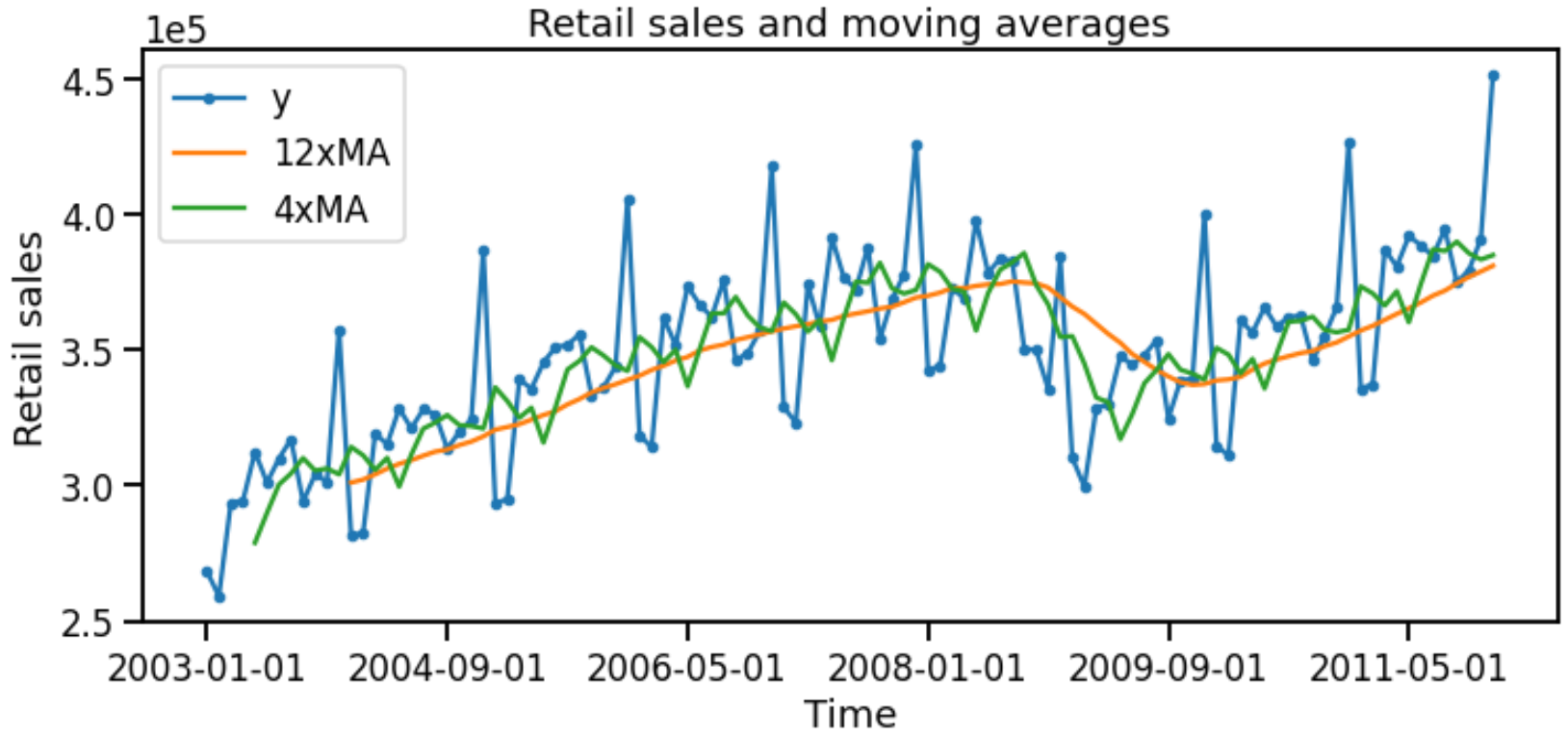
Motivation for weighted window functions

- We could use a shorter window but this would increase the variance of the new window feature.



Motivation for weighted window functions

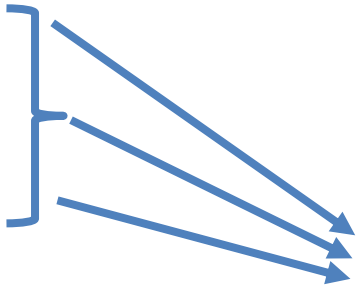
- A common solution is to assign weights to the window, more weight to recent observations, to compute, for example, a weighted mean.



Weighted rolling mean

- Let's look at the regular rolling mean.

Date	Sales
…	…
2020-02-12	23
2020-02-13	30
2020-02-14	35
2020-02-15	?



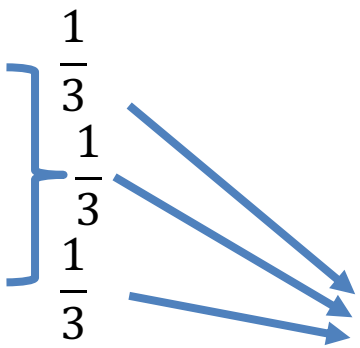
Rolling window (mean)
…
29.3

$$= \frac{1}{3}(35 + 30 + 23)$$

Weighted rolling mean

- Let's look at the regular rolling mean.

Date	Sales
…	…
2020-02-12	23
2020-02-13	30
2020-02-14	35
2020-02-15	?



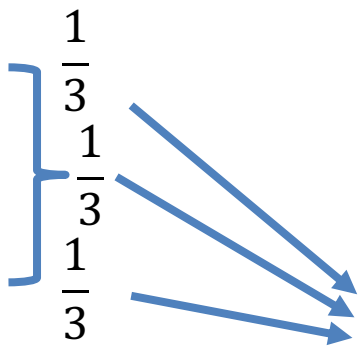
Rolling window (mean)
…
29.3

$$= \frac{1}{3}35 + \frac{1}{3}30 + \frac{1}{3}23$$

Weighted rolling mean

- What if we want to give more weight to recent time periods?

Date	Sales
…	…
2020-02-12	23
2020-02-13	30
2020-02-14	35
2020-02-15	?



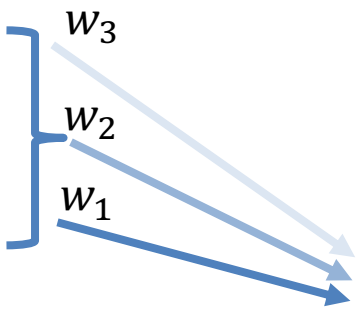
Rolling window (mean)
…
29.3

$$= \frac{1}{3}35 + \frac{1}{3}30 + \frac{1}{3}23$$

Weighted rolling mean

- We can specify weights to our window function to give a weighted average.

Date	Sales
…	…
2020-02-12	23
2020-02-13	30
2020-02-14	35
2020-02-15	?



Rolling window (mean)
…
x

$$= \frac{w_1 35 + w_2 30 + w_3 23}{\sum_i w_i}$$

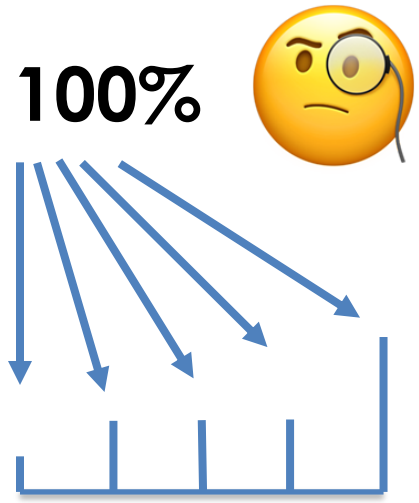
Other weighted statistics

- Each statistic (e.g., the mean) has a formula for a weighted version (e.g., weighted mean).

	Unweighted	Weighted
Mean	$\hat{\mu} = \frac{\sum_{i=1}^{N} x_i}{N}$	$\hat{\mu}_w = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$
Variance	$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} (x_i - \hat{\mu})^2}{N}$	$\hat{\sigma}_w{}^2 = \frac{\sum_{i=1}^{N} w_i (x_i - \hat{\mu}_w)^2}{\sum_{i=1}^{N} w_i}$

How to pick the weights?

- We can think of starting with 100% weight.
  We're spreading this over the window.
- How do we spread the weight?
  - Domain knowledge.
  - Linear & exponential weights.

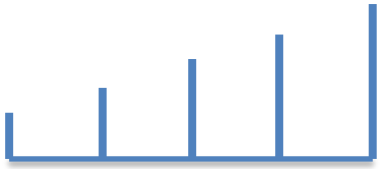


[10%, 20%, 20%, 20%, 30%]

How to pick the weights?

- We can think of starting with 100% weight. We're spreading this over the window.
- How do we spread the weight?
  - Domain knowledge.
  - Linear & exponential weights.

Linear



[6.5%, 13%, 19.5%, 26%, 32.5%]

Weight decays linearly

How to pick the weights?

- We can think of starting with 100% weight.
  We're spreading this over the window.
- How do we spread the weight?
  - Domain knowledge.
  - Linear & exponential weights.

Exponential

[3.2%, 6.5%, 13%, 26%, 52%]

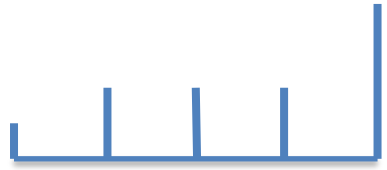
Weight decays exponentially

Parameter:
Rate of decay, α

How to pick the weights?

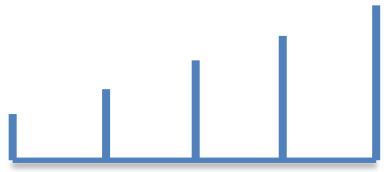
- We can think of starting with 100% weight. We're spreading this over the window.
- How do we spread the weight?
  - Domain knowledge.
  - Linear & exponential weights.
  - Try multiple weighting schemes and test performance (computationally prohibitive).
- The lack of a principled way to select weights is a downside of this approach.

Custom

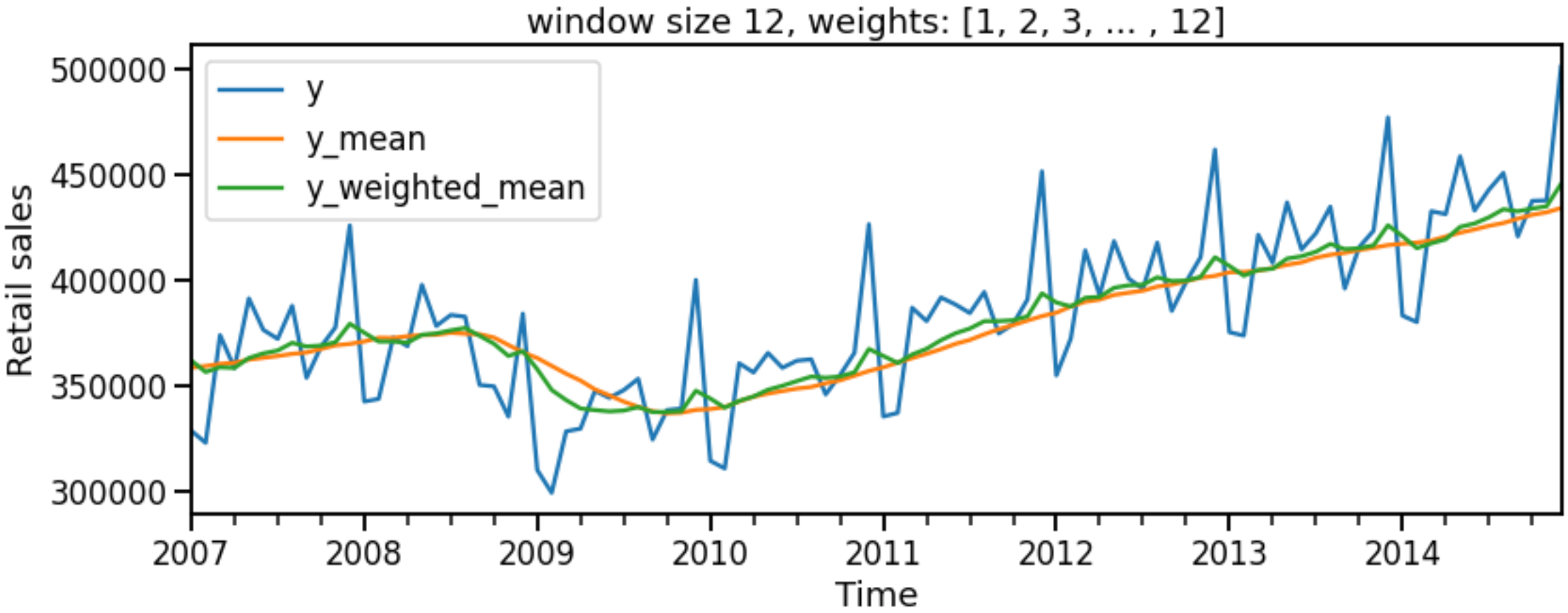


Exponential

Linear



Example: Retail sales with linear weights

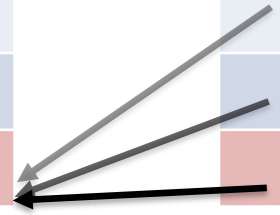


The weighted mean was more sensitive to the change in the trend between 2008 and 2010.

Aside: Relation to distributed lags

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

Ad spend	Ad spend Lag 1	Ad spend Lag 2
100	NaN	NaN
80	100	NaN
120	80	100
10	120	80



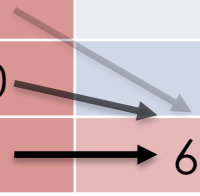
$$w_0 > w_1 > w_2$$

Distributed lags can result in many lag features. We also know we want more weight to be given to a recent lag relative to larger lags.

Aside: Relation to distributed lags

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

Ad spend	Ad Spend window
100	
80	
120	
10	65



Creating a weighted window feature allows us to condense the same logic into a single feature, with the downside that we must manually specify weights.

Summary

Weighting our window features can make the feature more sensitive to recent changes.

There are multiple ways that the weights can be defined.

There is no easy way to pick the “best” weights. Instead we can use heuristics, domain knowledge, and trial & error.

References

Simple exponential smoothing
https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm

https://otexts.com/fpp3/ses.html

Exponentially weighted moving averages
https://towardsdatascience.com/time-series-from-scratch-exponentially-weighted-moving-averages-ewma-theory-and-implementation-607661d574fe

https://web.archive.org/web/20100329135531/http://lorien.ncl.ac.uk/ming/filter/filewma.htm

https://www.investopedia.com/terms/e/ema.asp#toc-what-is-an-exponential-moving-average-ema

Linear weighted moving averages
https://www.investopedia.com/terms/l/linearlyweightedmovingaverage.asp

https://medium.com/the-investors-handbook/the-linear-weighted-moving-average-d70014c6d015