

Target encoding

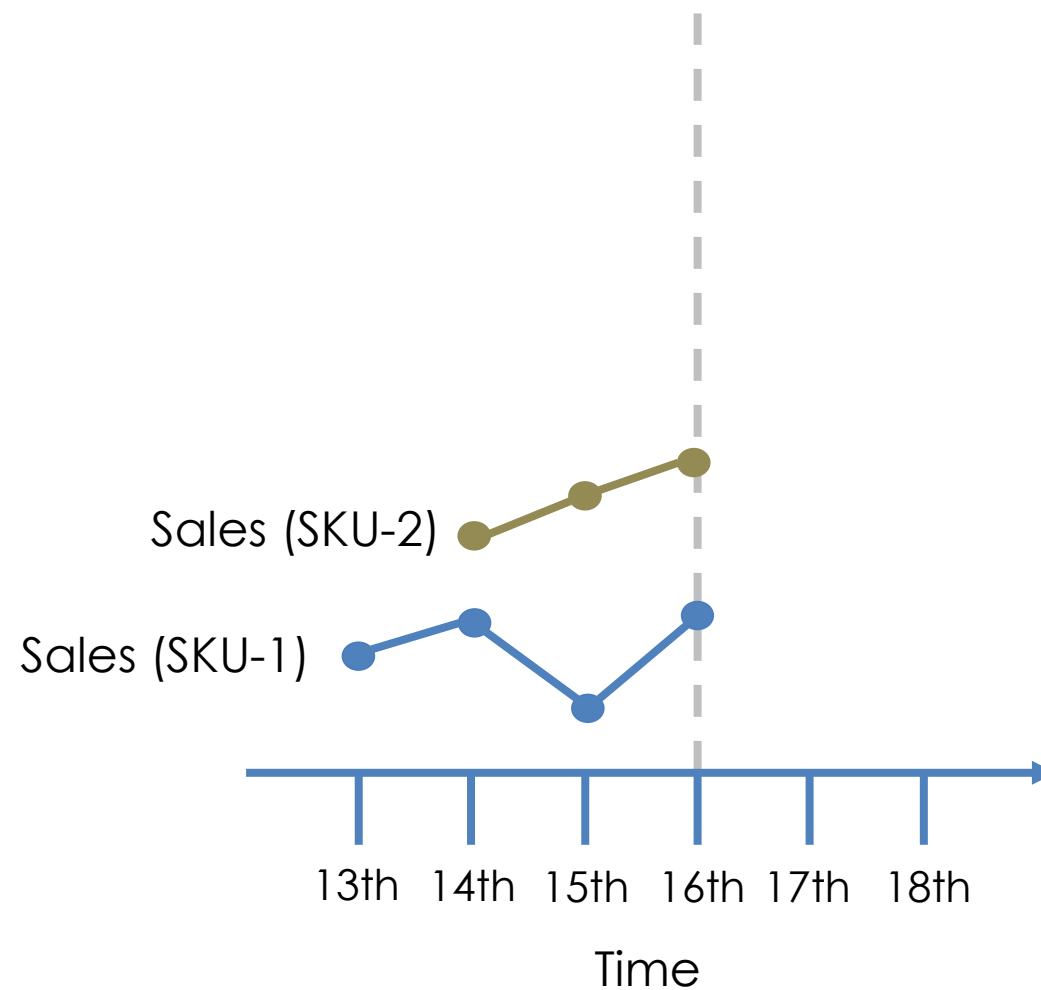
Method and
considerations

Target encoding

- Target encoding or mean encoding consists in replacing the category by the average target value for that category.
- Learns parameters from data.
- Careful with look-ahead bias.

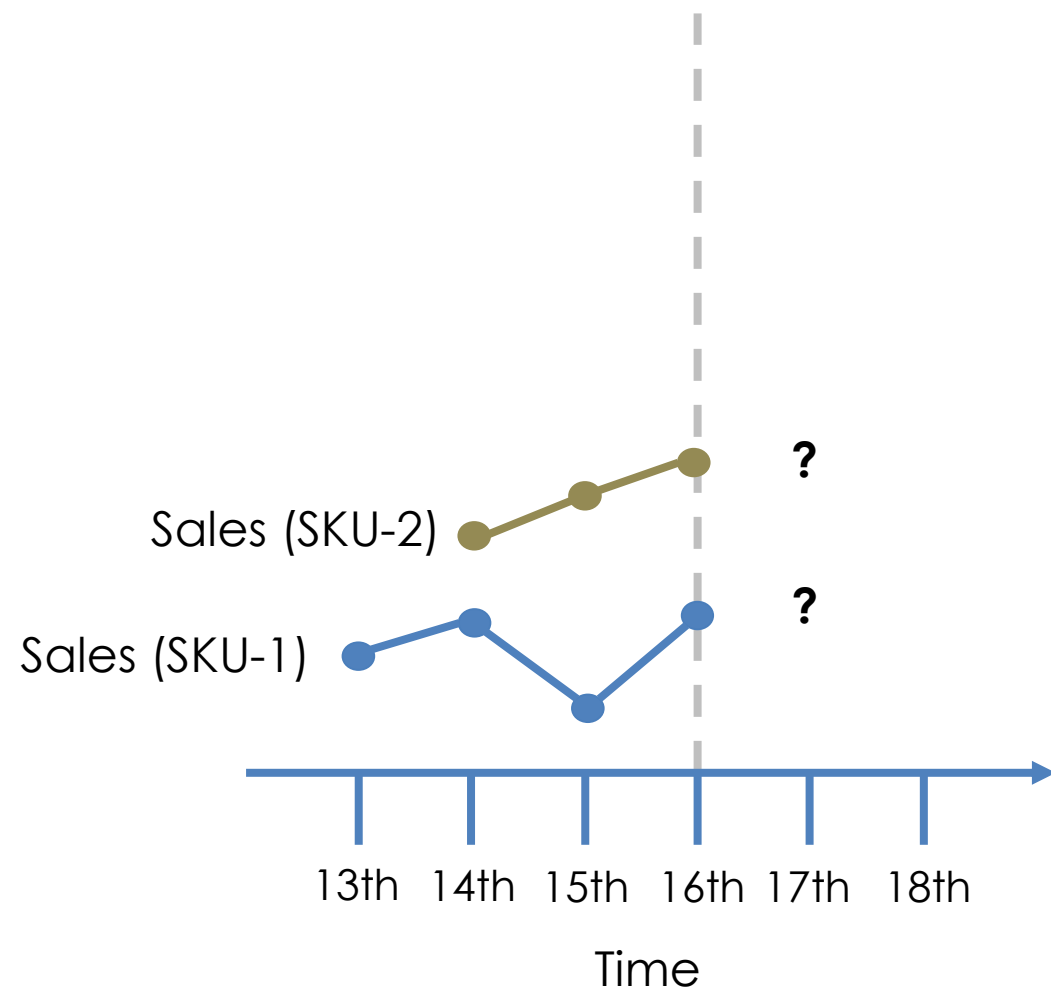
An example

Time	Product ID	Sales
...
2020-02-13	SKU-1	30
2020-02-14	SKU-1	32
2020-02-15	SKU-1	25
2020-02-16	SKU-1	34
...
2020-02-14	SKU-2	32
2020-02-15	SKU-2	21
2020-02-16	SKU-2	25



An example

Time	Product ID	Sales
...
2020-02-13	SKU-1	30
2020-02-14	SKU-1	32
2020-02-15	SKU-1	25
2020-02-16	SKU-1	34
2020-02-17	SKU-1	?
...
2020-02-14	SKU-2	32
2020-02-15	SKU-2	21
2020-02-16	SKU-2	15
2020-02-17	SKU-2	?



Static features

Time	Product ID
...	...
2020-02-13	SKU-1
2020-02-14	SKU-1
2020-02-15	SKU-1
2020-02-16	SKU-1
2020-02-17	SKU-1
...	...
2020-02-14	SKU-2
2020-02-15	SKU-2
2020-02-16	SKU-2
2020-02-17	SKU-2

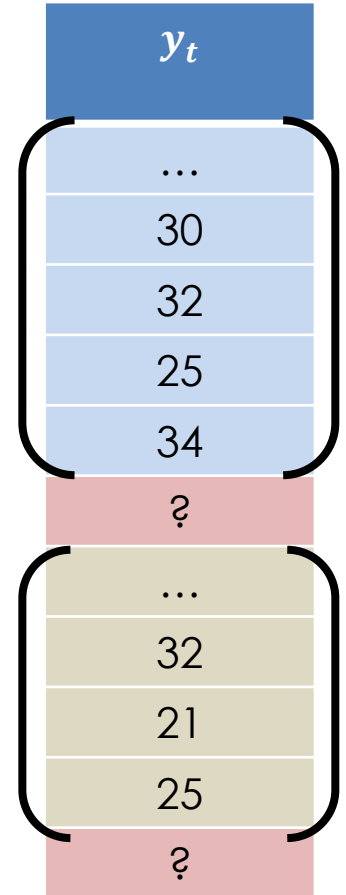
y_t
...
30
32
25
34
?
...
32
21
25
?

Static features: Target encoding

Time	Product ID
...	...
2020-02-13	SKU-1
2020-02-14	SKU-1
2020-02-15	SKU-1
2020-02-16	SKU-1
2020-02-17	SKU-1
...	...
2020-02-14	SKU-2
2020-02-15	SKU-2
2020-02-16	SKU-2
2020-02-17	SKU-2

SKU-1 = mean

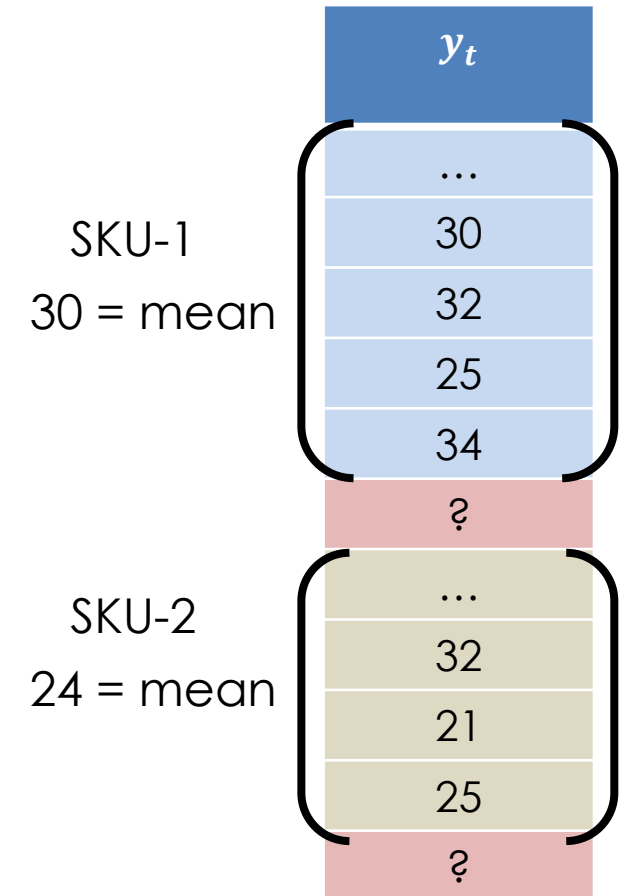
SKU-2 = mean



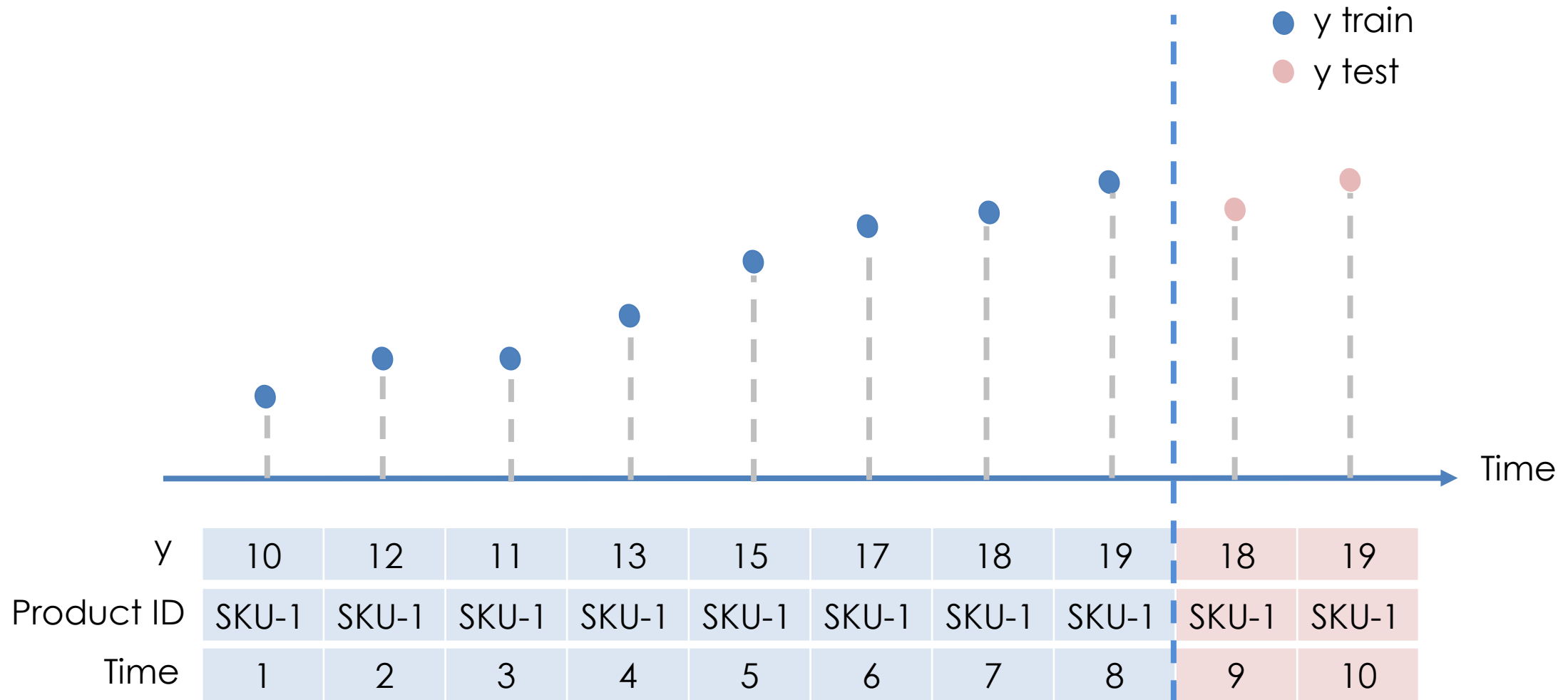
Static features: Target encoding

Time	Product ID (encoded)
...	...
2020-02-13	30
2020-02-14	30
2020-02-15	30
2020-02-16	30
2020-02-17	30
...	...
2020-02-14	24
2020-02-15	24
2020-02-16	24
2020-02-17	24

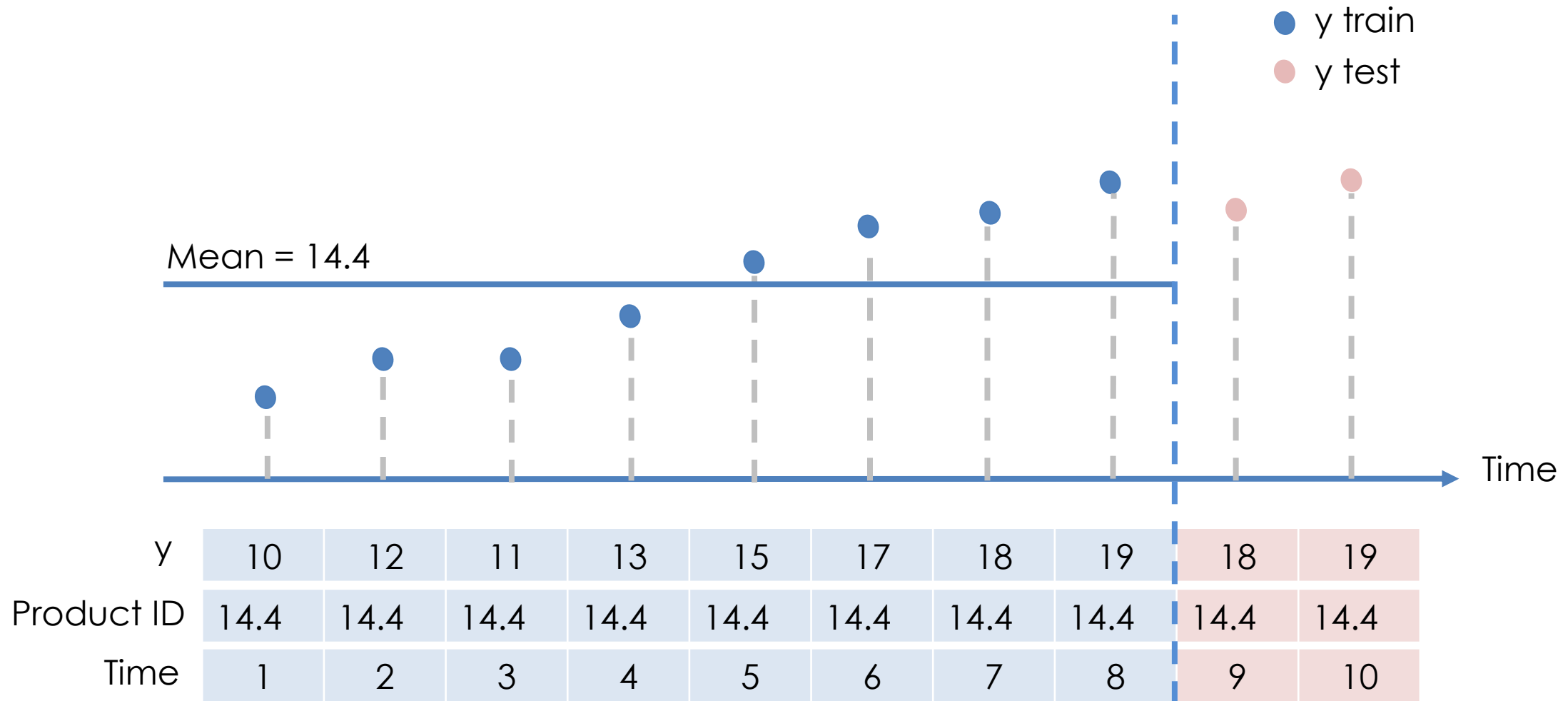
- No leakage between **train** and **test set**.
- But, the **target is leaked** from future values to past values in training set.
- Beware of overfitting.



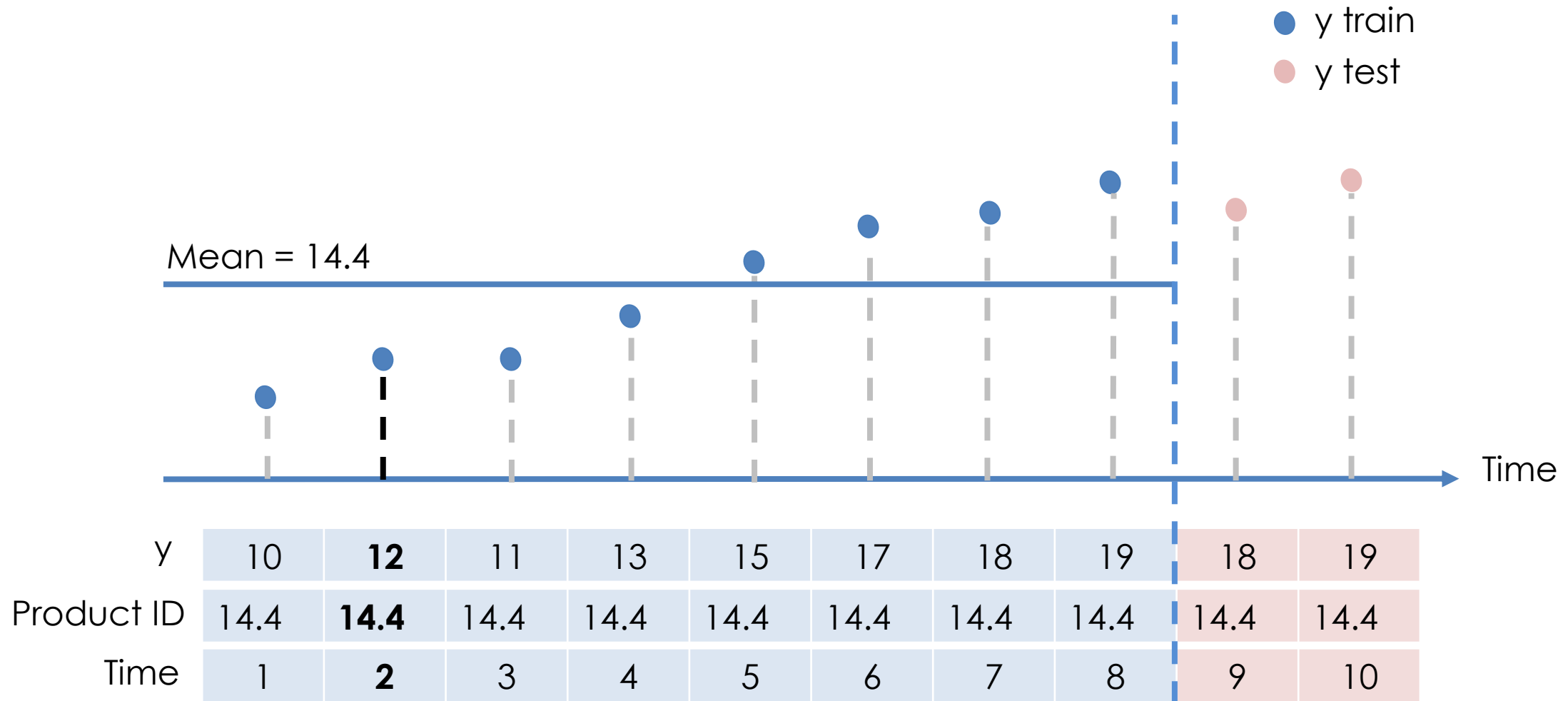
Static features: Target encoding



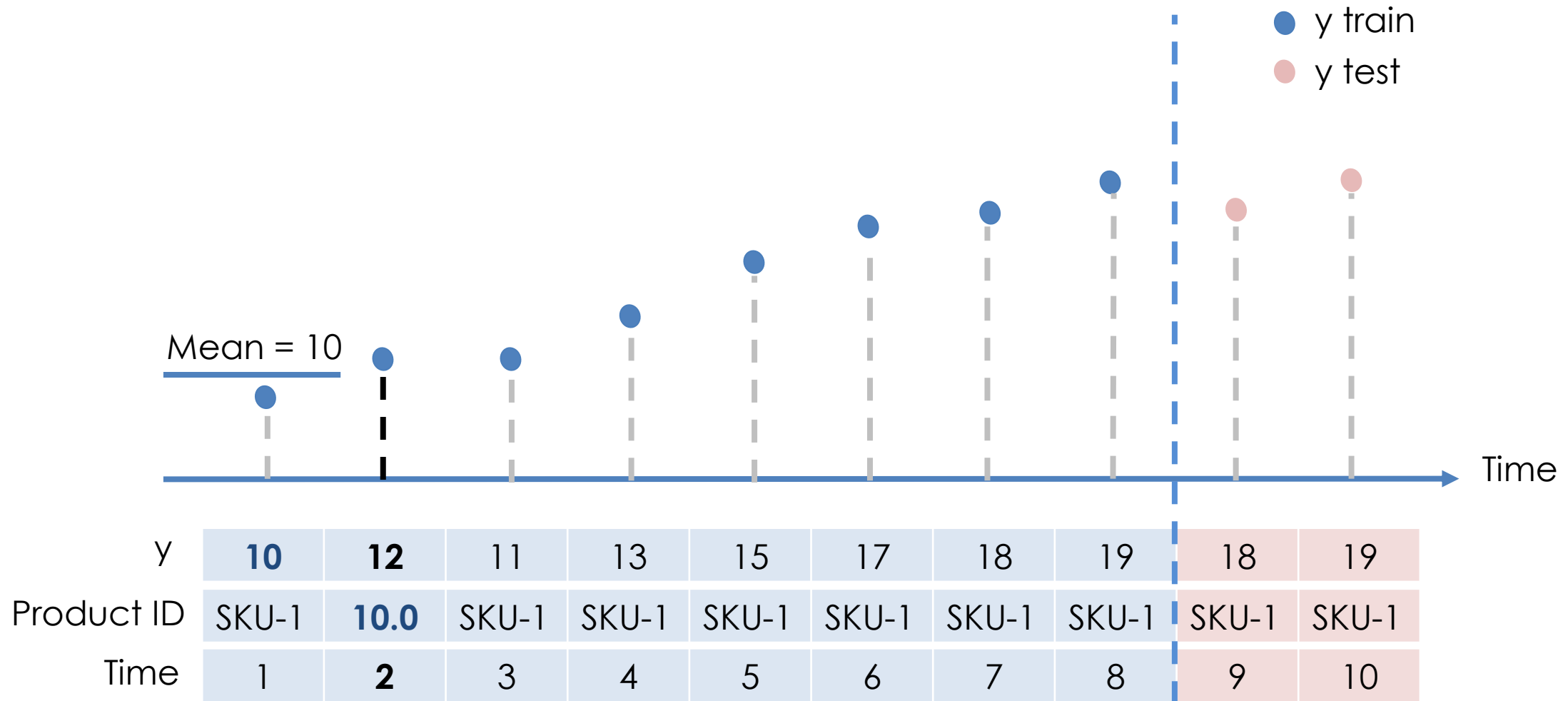
Static features: Target encoding



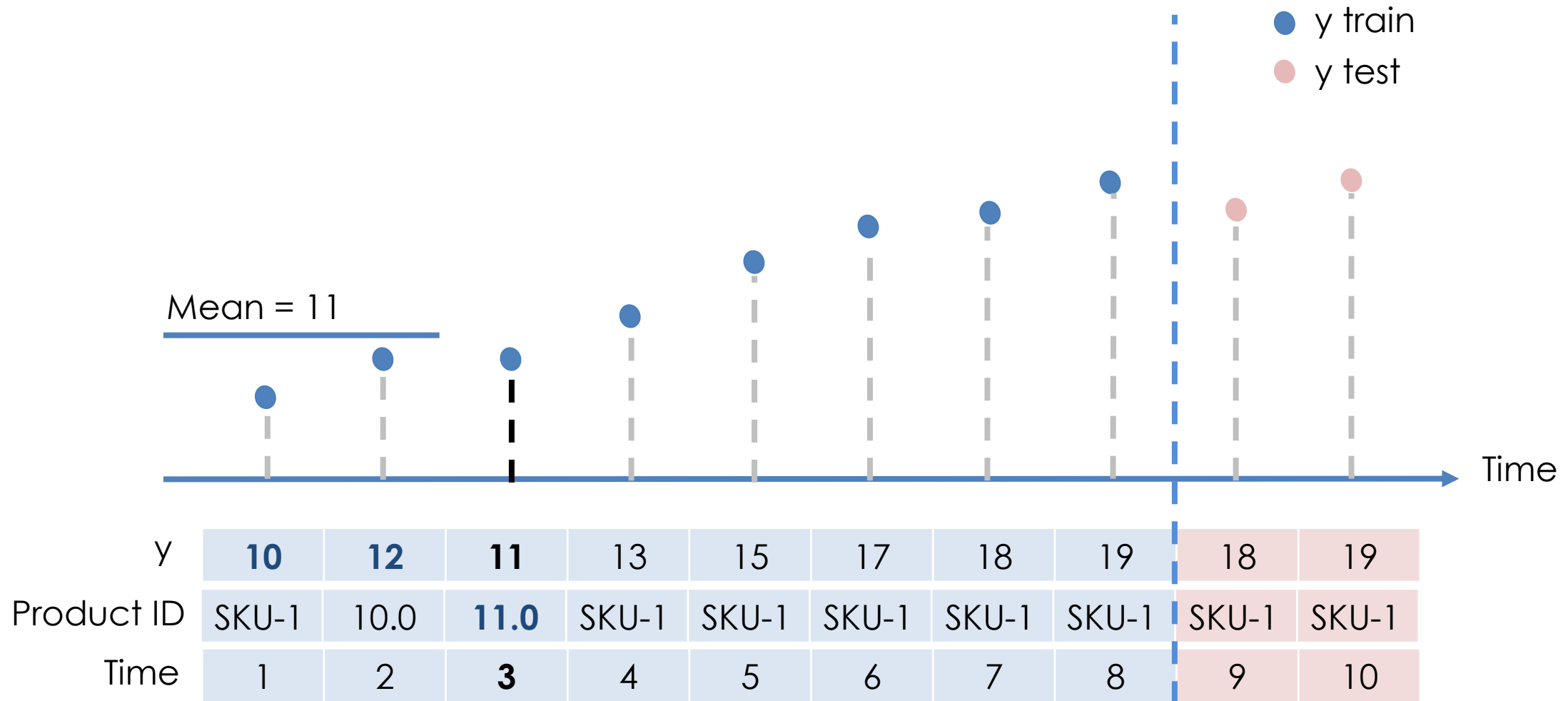
Static features: Target encoding



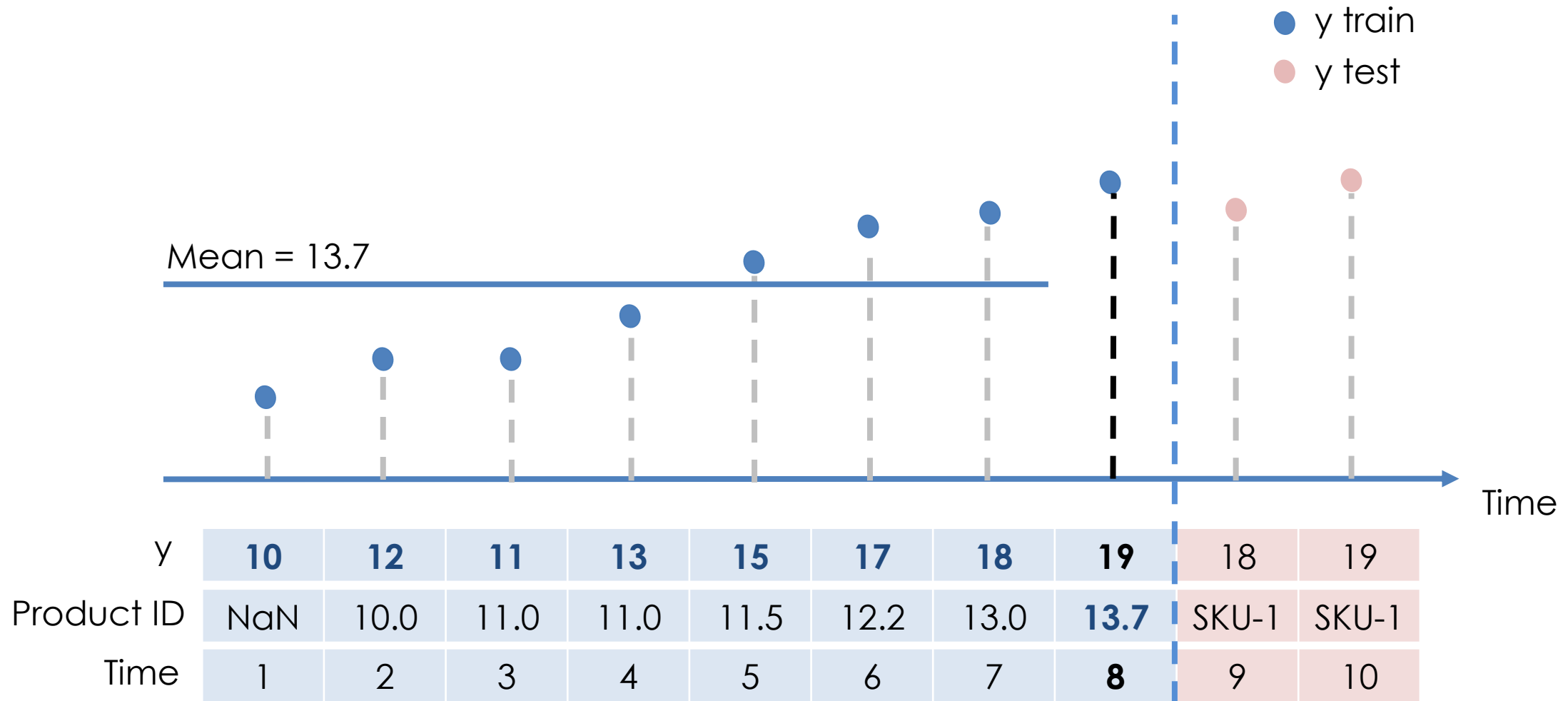
Static features: Target encoding



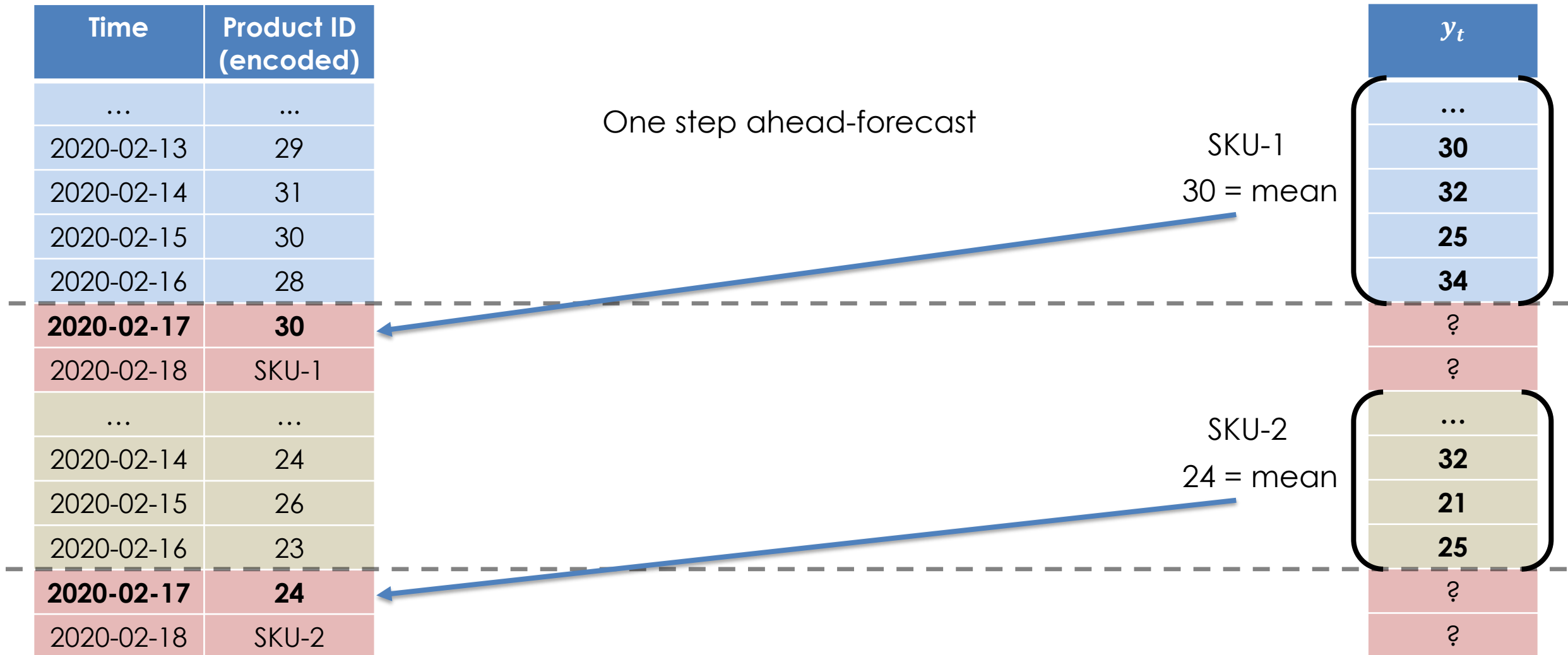
Static features: Target encoding



Static features: Target encoding



Static features: Target encoding



Static features: Target encoding

Time	Product ID (encoded)
...	...
2020-02-13	29
2020-02-14	31
2020-02-15	30
2020-02-16	28

For **recursive forecasting**, we need to dynamically compute the encoding at predict time.

SKU-1
30 = mean

y_t
...
30
32
25
34

2020-02-17	30
2020-02-18	SKU-1
...	...
2020-02-14	24
2020-02-15	26
2020-02-16	23
2020-02-17	24
2020-02-18	SKU-2

SKU-2
24 = mean

...
32
21
25
?
?

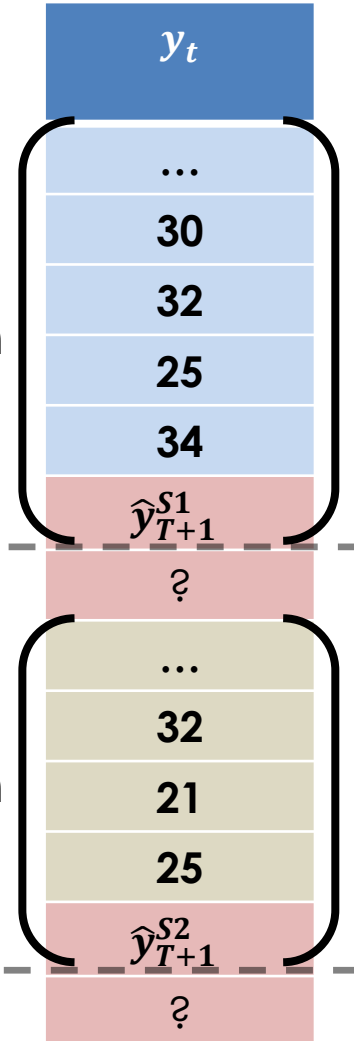
Static features: Target encoding

For **recursive forecasting**, we need to dynamically compute the encoding at predict time.

Time	Product ID (encoded)
...	...
2020-02-13	29
2020-02-14	31
2020-02-15	30
2020-02-16	28
2020-02-17	30
2020-02-18	SKU-1
...	...
2020-02-14	24
2020-02-15	26
2020-02-16	23
2020-02-17	24
2020-02-18	SKU-2

SKU-1 = mean

SKU-2 = mean



Summary

Target encoding consists in replacing categories with the mean target value.

Risk of leaking future values of the target into the features.