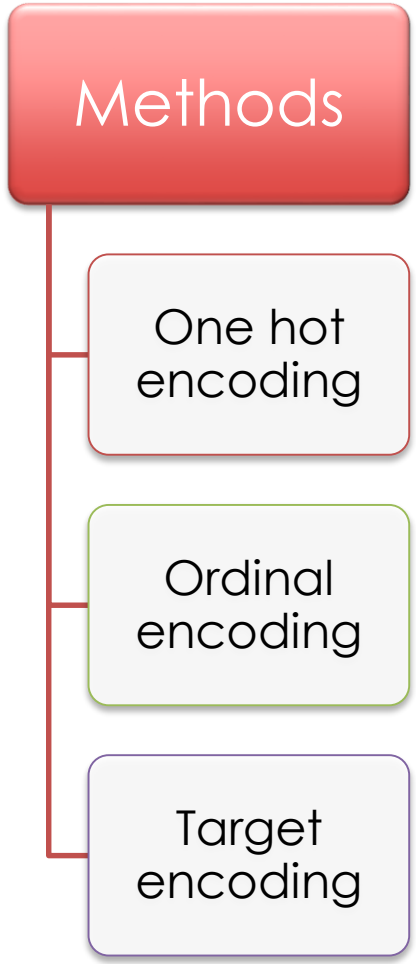
Summary

Categorical encoding

Encoding methods



One-hot encoding

Country
UK
Spain
Germany
UK

UK	Spain	Germany
1	0	0
0	1	0
0	0	1
1	0	0

- Suitable for linear models (k-1 dummies)
- Can create k dummies for trees
- Expand the feature space if multiple static features, or if they are highly cardinal

Ordinal encoding

Country
UK
Spain
Germany
UK

Country
0
1
2
0

- Suitable for non-linear models.

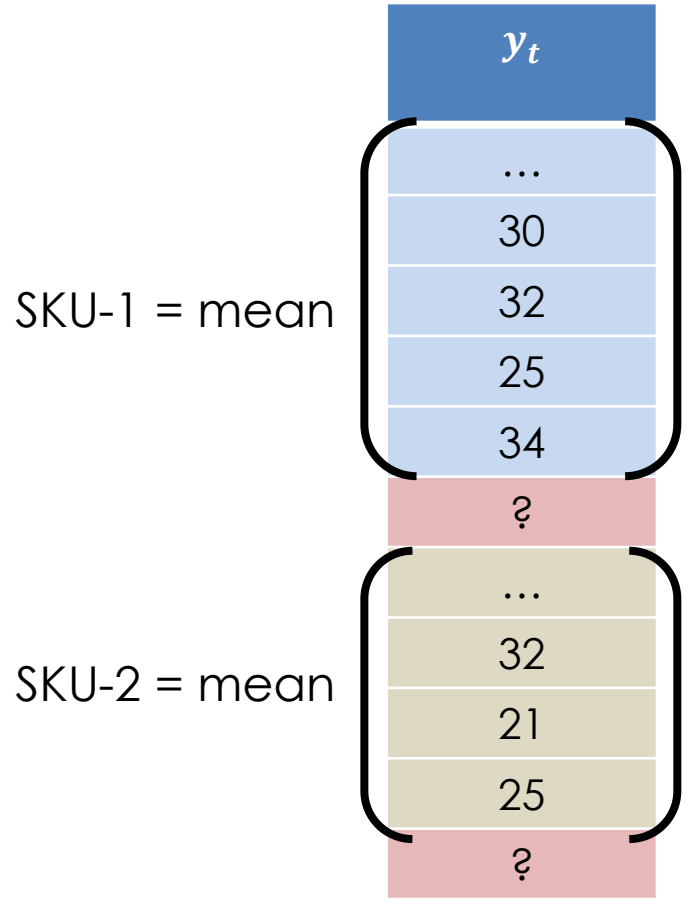
Common encoding methods

Feature-Engine

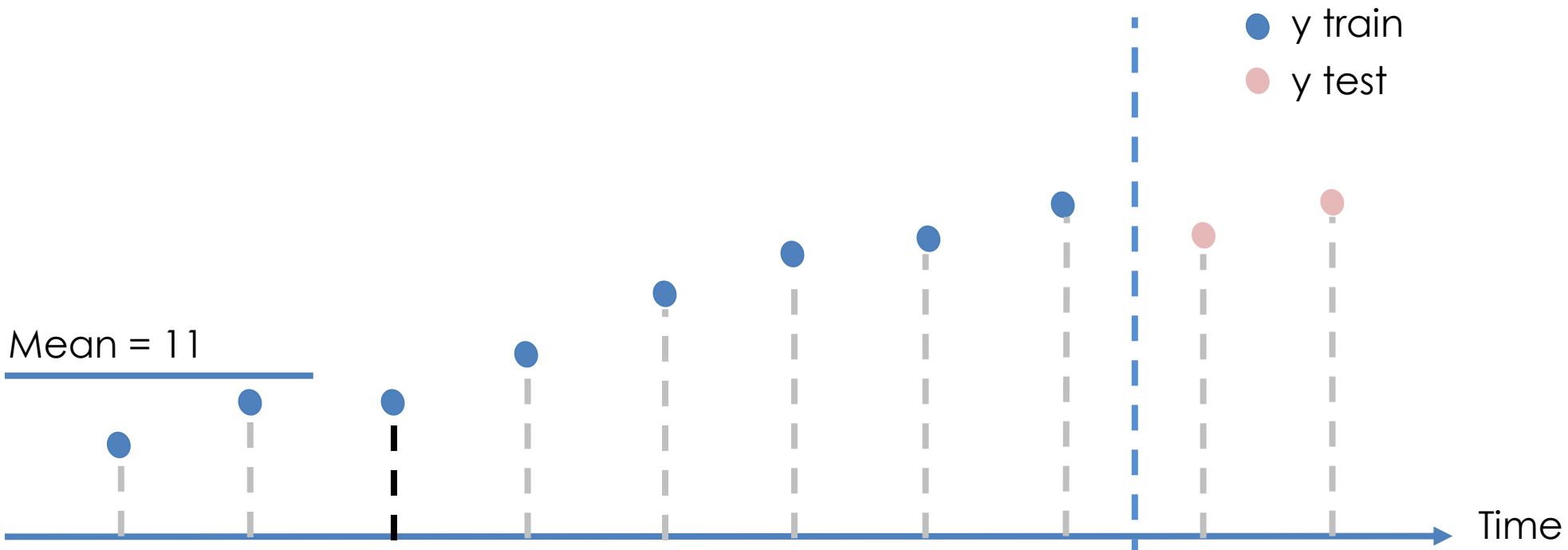
Category Encoders

Target encoding: train / test

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



Target encoding: expanding windows



y	**10**	**12**	**11**	13	15	17	18	19	18	19
Product ID	SKU-1	10.0	**11.0**	SKU-1	SKU-1	SKU-1	SKU-1	SKU-1	SKU-1	SKU-1
Time	1	2	**3**	4	5	6	7	8	9	10

Target encoding: expanding windows

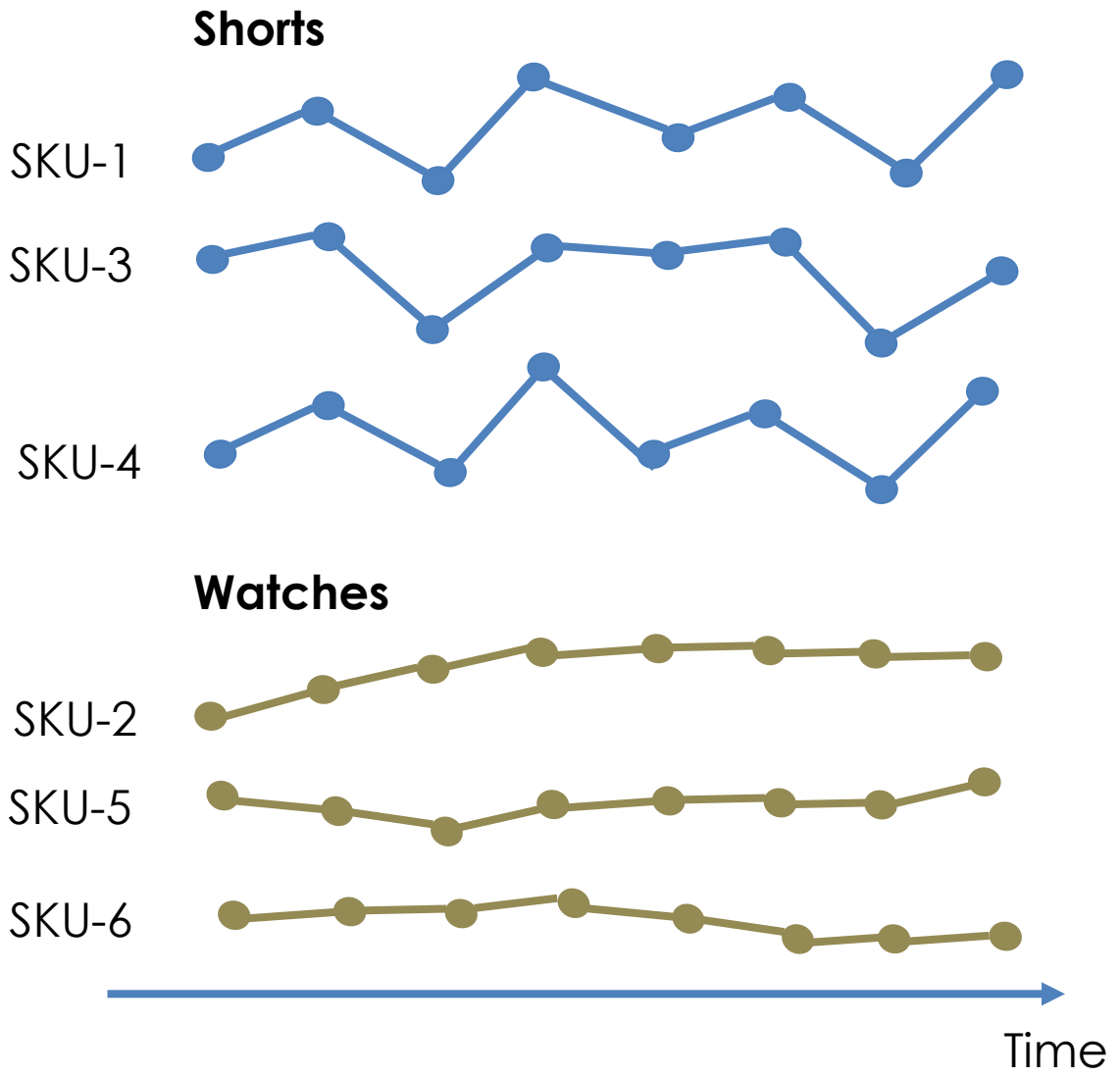
Time	Product ID	y_t (expanding window)
…	…	…
2020-02-13	SKU-1	…
2020-02-14	SKU-1	30
2020-02-15	SKU-1	(30+32)/2
2020-02-16	SKU-1	(30+32+25)/3
2020-02-17	SKU-1	(30+32+25+34)/4
…	…	…
2020-02-14	SKU-2	…
2020-02-15	SKU-2	32
2020-02-16	SKU-2	(32+21)/2
2020-02-17	SKU-2	(32+21+25)/3

The target encoding is in essence creating an expanding window feature of the target variable.

y_t
…
30
32
25
34
?
…
32
21
25
?

Multiple static features

Time	Product ID	Product category
…	…	…
2020-02-13	SKU-1	Shorts
2020-02-14	SKU-1	Shorts
2020-02-15	SKU-1	Shorts
2020-02-16	SKU-1	Shorts
2020-02-17	SKU-1	Shorts
…	…	…
2020-02-14	SKU-2	Watches
2020-02-15	SKU-2	Watches
2020-02-16	SKU-2	Watches
2020-02-17	SKU-2	Watches



Multiple static features

Expanding windows for shorts different from expanding windows for product ids.