Train In Data

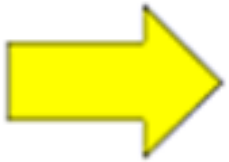
# Mean or target encoding

# Mean encoding: definition

- Mean encoding implies replacing the category by the average target value for that category

# Mean encoding: example

| Color |
| --- |
| Red |
| Red |
| Yellow |
| Green |
| Yellow |



| Target | Color |
| --- | --- |
| 0 | 0.5 |
| 1 | 0.5 |
| 1 | 1 |
| 0 | 0 |
| 1 | 1 |

# Mean encoding: Advantages

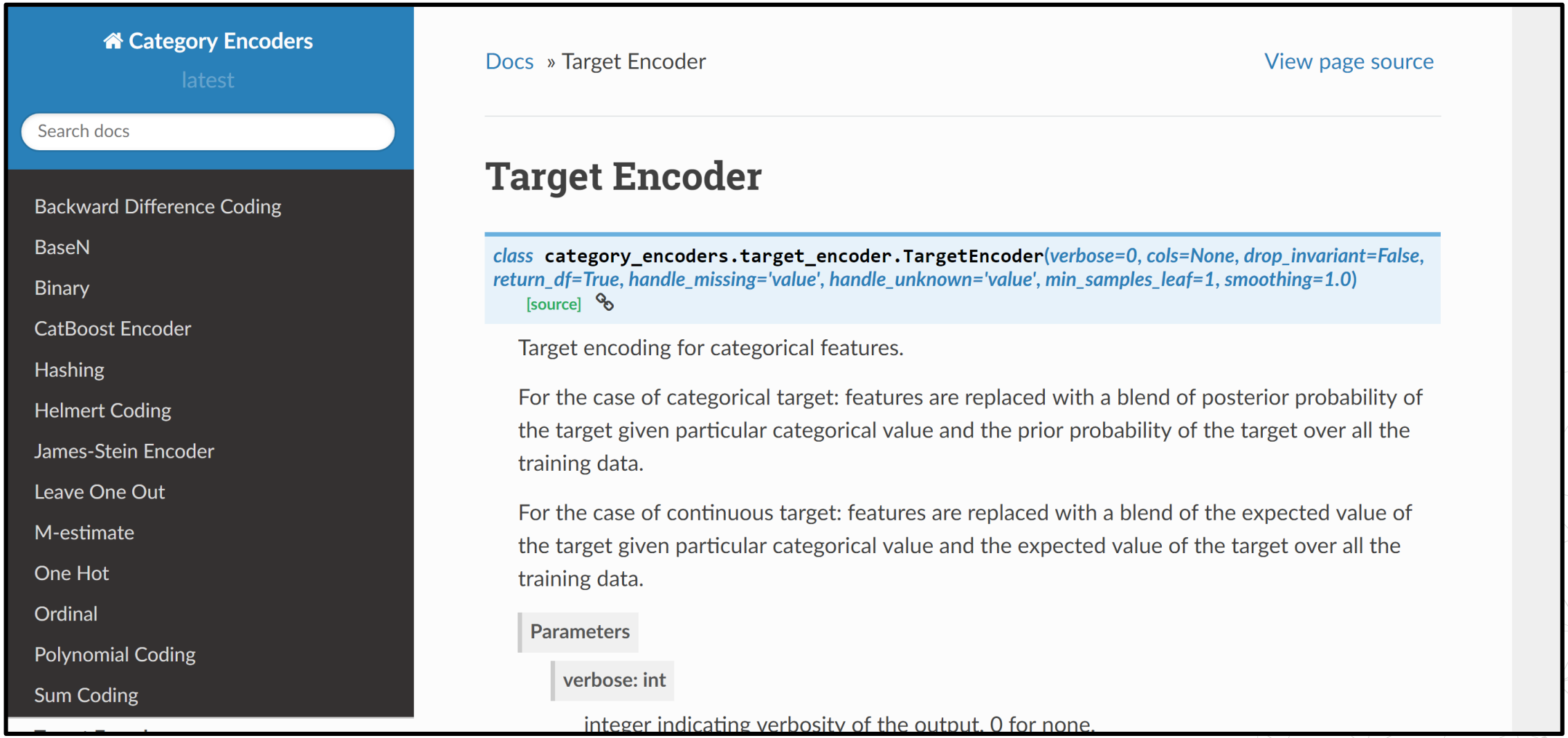
- Straightforward to implement
- Does not expand the feature space
- Creates monotonic relationship between categories and target

# Mean encoding: Limitations

- May lead to over-fitting

- If 2 categories show the same mean of target, they will be replaced by the same number => potential loss of value

# Mean encoding with Category Encoders



Category Encoders

latest

Search docs

- Backward Difference Coding
- BaseN
- Binary
- CatBoost Encoder
- Hashing
- Helmert Coding
- James-Stein Encoder
- Leave One Out
- M-estimate
- One Hot
- Ordinal
- Polynomial Coding
- Sum Coding

Docs » Target Encoder

View page source

## Target Encoder

*class* **category_encoders.target_encoder.TargetEncoder**(*verbose=0, cols=None, drop_invariant=False, return_df=True, handle_missing='value', handle_unknown='value', min_samples_leaf=1, smoothing=1.0*)

[source]

Target encoding for categorical features.

For the case of categorical target: features are replaced with a blend of posterior probability of the target given particular categorical value and the prior probability of the target over all the training data.

For the case of continuous target: features are replaced with a blend of the expected value of the target given particular categorical value and the expected value of the target over all the training data.

**Parameters**

**verbose: int**

integer indicating verbosity of the output. 0 for none.

Train In Data

# THANK YOU

www.trainindata.com