



Weight of Evidence

Weight of evidence: definition

$$\text{WOE} = \ln \left(\frac{\text{Distribution of Goods}}{\text{Distribution of Bads}} \right)$$

Weight of Evidence (WoE) was developed primarily for the credit and financial industries to help build more predictive models to evaluate the risk of loan default.


That is, to predict how likely the money lent to a person or institution is to be lost.

Weight of evidence: definition

$$\text{WOE} = \ln \left(\frac{\text{Distribution of Goods}}{\text{Distribution of Bads}} \right)$$

- Usually in finance $\text{WoE} = \ln (p(0) / p(1))$
 - Where $p(1)$ is the probability of default
- In the course and in Feature-engine: $\text{WoE} = \ln (p(1) / p(0))$
 - Just changes the sign of the output

Weight of evidence: example

	survived	died		ratio
cabin				
A	0.411765	0.588235	$\ln\left(\frac{p(1)}{p(0)}\right)$ 	-0.356675
B	0.738095	0.261905		1.036092
C	0.600000	0.400000		0.405465
D	0.696970	0.303030		0.832909
E	0.700000	0.300000		0.847298
F	0.769231	0.230769		1.203973
G	0.750000	0.250000		1.098612
n	0.292199	0.707801		-0.884730



Weight of evidence: Advantages

- Creates a monotonic relationship between the target and the variables.
- It orders the categories on a "logistic" scale which is natural for logistic regression
- The transformed variables can then be compared because they are on the same scale.
 - Therefore, it is possible to determine which one is more predictive.

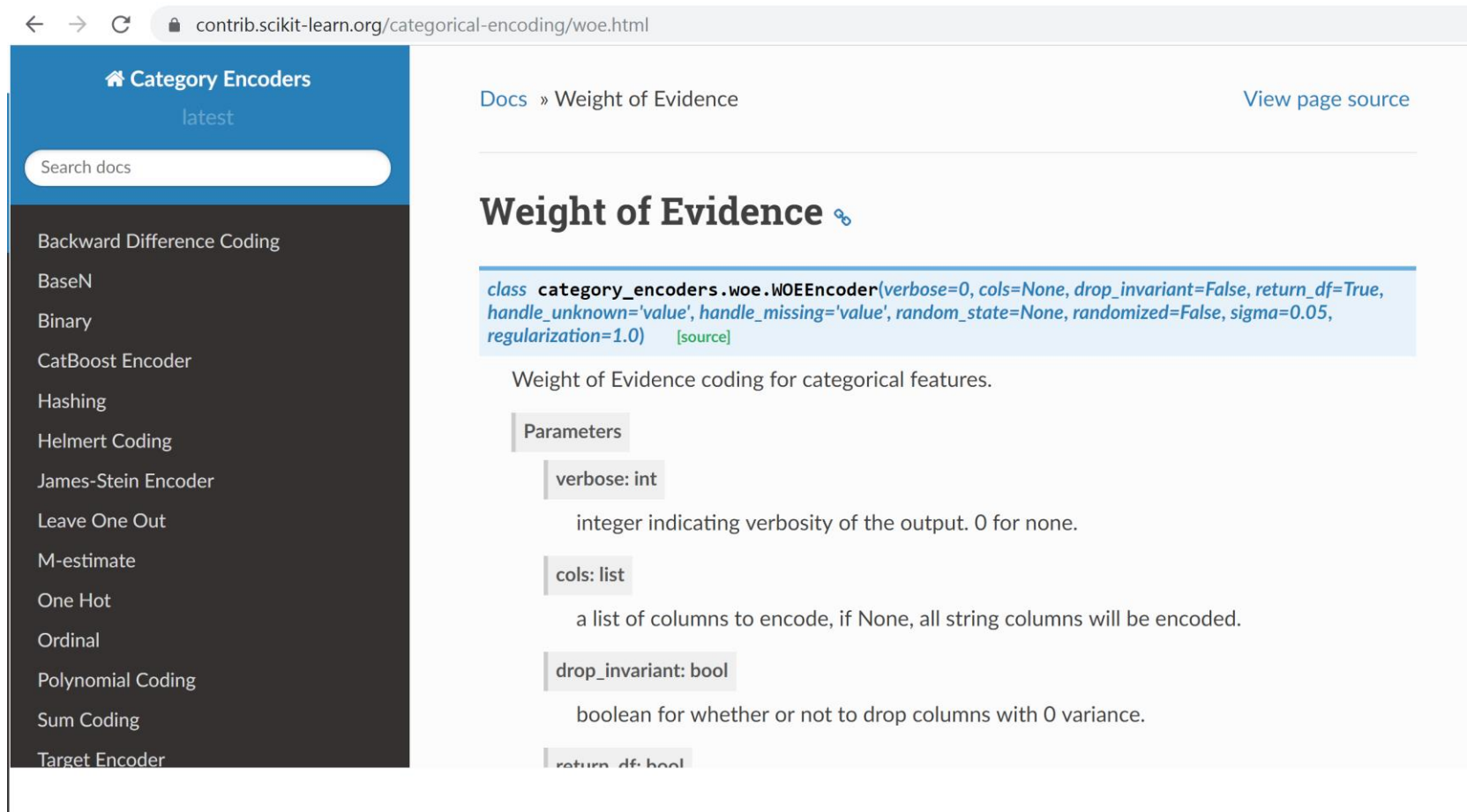


Weight of evidence: Limitations

- May lead to over-fitting
- Not defined when the denominator is 0



Weight of evidence with Category Encoders



The screenshot shows the Scikit-Learn documentation page for Category Encoders. The left sidebar lists various encoders, and the main content area details the Weight of Evidence (WOE) encoder.

← → ↻ contrib.scikit-learn.org/categorical-encoding/woe.html

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
- Target Encoder

Docs » Weight of Evidence [View page source](#)

Weight of Evidence 🔗

```
class category_encoders.woe.WOEEncoder(verbose=0, cols=None, drop_invariant=False, return_df=True, handle_unknown='value', handle_missing='value', random_state=None, randomized=False, sigma=0.05, regularization=1.0) [source]
```

Weight of Evidence coding for categorical features.

Parameters

- verbose: int**
integer indicating verbosity of the output. 0 for none.
- cols: list**
a list of columns to encode, if None, all string columns will be encoded.
- drop_invariant: bool**
boolean for whether or not to drop columns with 0 variance.
- return_df: bool**

THANK YOU

www.trainindata.com