Train In Data

Weight of Evidence

Weight of evidence: definition

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

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

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

- Usually in finance WoE = ln (p(0) / p(1))
  - Where p(1) is the probability of default

- In the course and in Feature-engine: WoE = ln (p(1) / p(0))
  - Just changes the sign of the output

Weight of evidence: example

cabin	survived	died	ratio
A	0.411765	0.588235	-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

$$\ln\left(\frac{p(1)}{p(0)}\right)$$

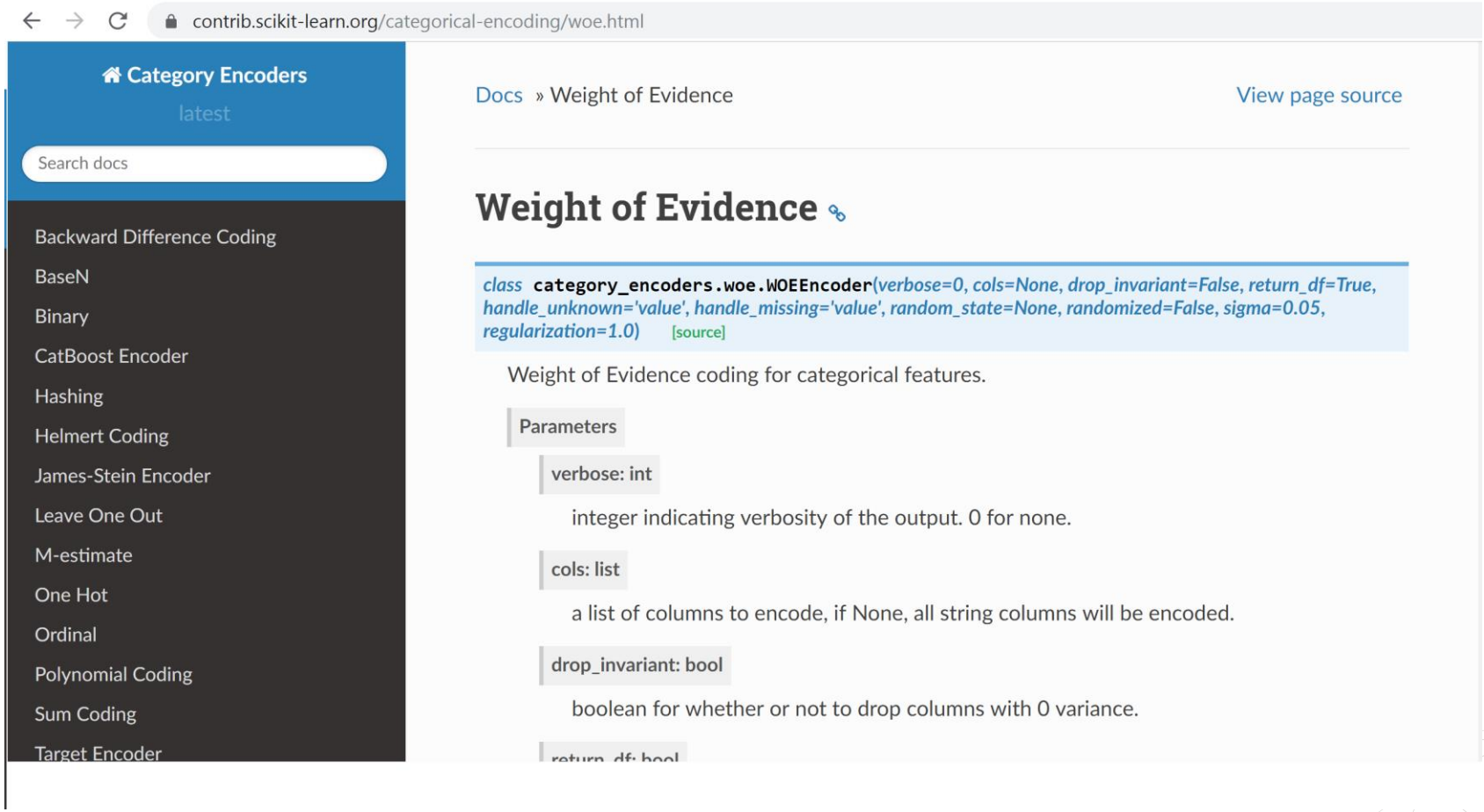
Weight of evidence: Advantages

- 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



Train In Data

THANK YOU

www.trainindata.com