Train In Data

Monotonic Encoding

Categorical encoding - definition

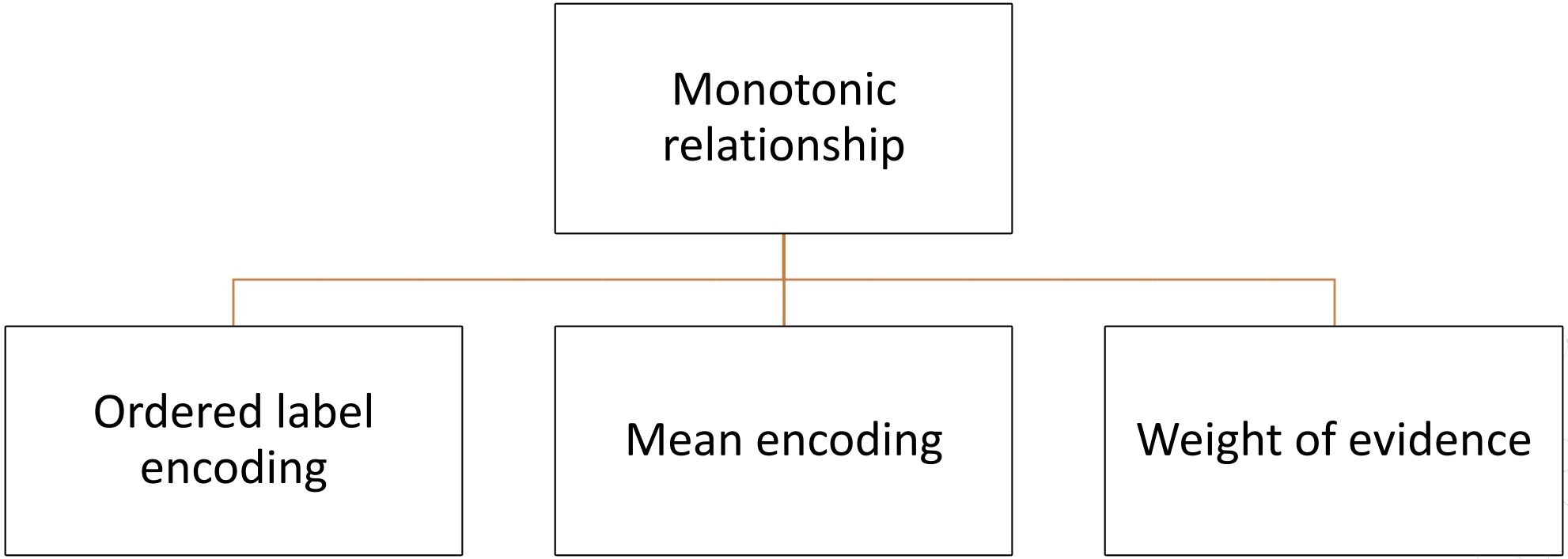
Categorical encoding refers to replacing the category strings by a numerical representation.

Categorical encoding - goal

- To produce variables that can be used to train machine learning models.

- To build predictive features from categories.

Monotonic encoding - methods

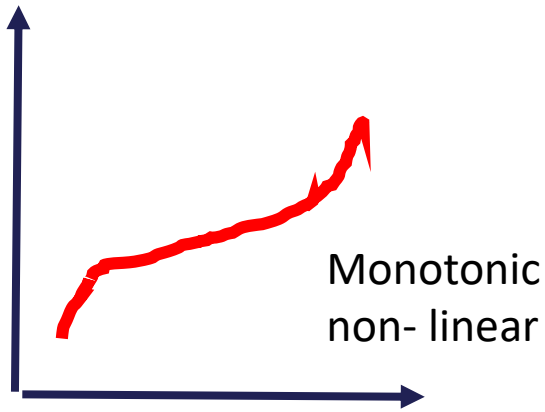


Monotonic relationship

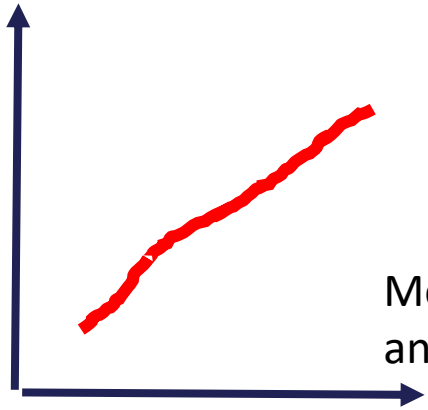
What is a monotonic relationship?

- When variable increases and the target increases. Or,
- When variable increases and the target decreases.

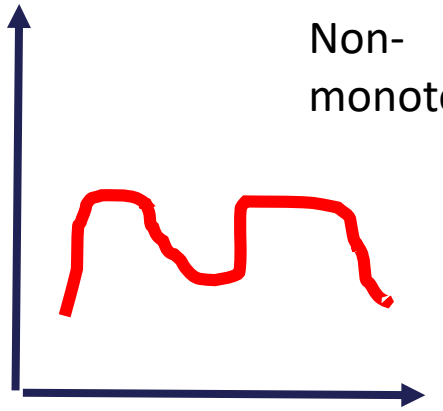
Monotonic relationship - examples



Monotonic non- linear



Monotonic and linear



Non-monotonic

Non-monotonic

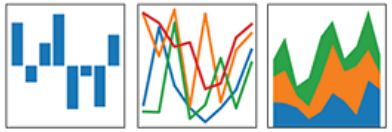
Monotonic relationship – why?

- Improve the performance of linear models
- May improve the performance of tree based models
  - Creates shallower trees

Monotonic encoding and Python

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Feature-Engine

Category Encoders

Content

For each lecture:

- Presentation and video
- Accompanying Jupyter notebook
  - Explanation of the technique
  - Implementation in **pandas**
  - Implementation in **Feature-engine**
  - Implementation in **Category encoders**

Final Summary

- Final lecture comparing the performance of the different categorical encoding techniques with different machine learning models.

- Additional reading resources.

Train In Data

THANK YOU

www.trainindata.com