Feature Engineering, Dimensionality Reduction Basics

Content

Feature engineering

- Steps
- Types
  - -Min Max, Normal, Binary, Range, Standardize

Dimensionality reduction

- PCA, t-SNE, Singular Value Decomposition, High correlation, Low Variance, Using ML model

Feature engineering: Steps

1. Brainstorming or Testing features
2. Deciding what features to create.
3. Creating features
4. Checking how the features work with your model
5. Improving your features if needed
6. Go back to brainstorming/creating more features until the work is done

Feature engineering: Types

1. Min Max
2. Binary
3. Binning (Discretization)
4. Standardize
5. Square / cube root
6. Logarithmic
7. One-hot encoding
8. 0-1

Min-Max

1. Min Max

$$x_{new} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}}$$

Binary

2. Binary

$$x_{new} = \begin{cases} 0, & x_{old} < Specified\ value \\ 1, & x_{old} \geq Specified\ value \end{cases}$$

Bining

3. Binning (Discretize)

$$x_{new} = \begin{cases} 0 & k_1 < x_{old} \leq k_2 \\ 1 & k_2 < x_{old} \leq k_3 \\ 2 & k_3 < x_{old} \leq k_4 \end{cases}$$

Standardize

4. Standardize (Normal)

$$x_{new} = \frac{x_{old} - mean(x)}{SD(x)}$$

Root

5. Square / Cube root

$$x_{new} = \sqrt[n]{x}, n = 2,3, \ldots$$

Logarithmic

6. Logarithmic

$$x_{new} = log_k(x)$$

One-Hot Encoding

7. One-hot encoding

Occupation	X_Business	X_Salaried	X_Unemployed
Business	1	0	0
Salaried	0	1	0
Salaried	0	1	0
Unemployed	0	0	1
Business	1	0	0
Salaried	0	1	0
Unemployed	0	0	1
Unemployed	0	0	1

Zero-One

8. Zero-One

Gender	X_new
M	1
M	1
F	0
M	1
F	0

The End