Mathematical Transformations

Tuesday 14 May 2024 9:41 PM

→ Transform then scale

→ To improve model performance → transformations conform to model assumptions & which in turn amplifies model's predictive power.

- even out variance
- make feature more normal
- reduce the skew
- linearize the relationship b/w feature & target
- Reduce the impact of outliers

Parametric → assumptions about data models

→ Mathematical operations to modify the original data features in a way that enhances their representation for ML models.

→ These transformations can help in improving model accuracy, meeting algo assumptions.

Types
- Power → aims to transform to a symmetric distribution NOT Normal.
- Log transformations
- Complex

problems after
- Interpretation
- Finding the best transformation is tricky (Experiment)
- Additional step in the pipeline

1) Log transformation

large value → more compression

Use
- → Right skewed data [moderate] ≃ 1-3 only
- → data contains outliers [RobustScaler or LogMinMax/standardScaler]
- → Reduces Hetroskedasticity

Don't use
- -ve or 0 values
- normal or uniform distribution already
- Interpretation [unit → % change] ← Research

→ Invertible

2) Square Root transformation

Sqrt ≃ log

→ log is extreme version of sqrt in compression of larger values

→ mild right skewed data [0-1]

3) Square transformations

→ mild left skewed data [-1, 0]

→ linearize non-linear relationships
  ↳ polynomial features

4) Reciprocal Transformation

Don't use → left skewed with outlier close to 0

use → Strong Right skewed (x > 1) → y [0, 1]
  [> 4]
  ↳ linearize relationship b/w feature & target

→ Strong left skewed → reflect → shift (Now treat this a right skewed distribution)
  (-) → (MAX+1) - X

5) Box-Cox transform

$$y'(\lambda) = \begin{cases} \frac{y^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y) & \text{if } \lambda = 0 \end{cases}$$

$\lambda \rightarrow [-5 \text{ to } 5]$ MLE + optimization

use → handle skewness
  → handle non-normal data
  → Heteroscedasticity
  → general case to many other transformations

Don't use → -ve values
  → Interpretability
  → Data already normally distributed
  → Categorical Data

6) Yeo Johnson transformation

→ improvement on Box-cox can handle -ve values also.

Yeo–Johnson transformation [edit]

The Yeo–Johnson transformation[15] allows also for zero and negative values of y. λ can be any real number, where $\lambda = 1$ produces the identity transformation. The transformation law reads:

$$y_i^{(\lambda)} = \begin{cases} ((y_i + 1)^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \ln(y_i + 1) & \text{if } \lambda = 0, y \geq 0 \\ -((-y_i + 1)^{(2-\lambda)} - 1)/(2 - \lambda) & \text{if } \lambda \neq 2, y < 0 \\ -\ln(-y_i + 1) & \text{if } \lambda = 2, y < 0 \end{cases}$$