Feature Engineering & Selection

Eng Teong Cheah
MVP Visual Studio &
Development Technologies

Agenda

Using feature engineering
Using feature selection

Using feature engineering

Feature engineering

Feature engineering attempts to increase the predictive power of learning algorithms by creating features from raw data that help facilitate the learning process.

Feature engineering and selection are parts of the **Develop features** step of the TDSP.

Feature engineering

Feature engineering: This process attempts to create additional relevant features from the existing raw features in the data, and to increase the predictive power of the learning algorithm.

Feature selection: This process selects the key subset of original data features in an attempt to reduce the dimensionality if the training problem.

Feature engineering

Normally **feature engineering** is applied first to generate additional features, and then the **feature selection** step is performed to eliminate irrelevant, redundant, or highly correlated features.

Feature selection

Feature selection

Feature selection is an important tool in machine learning.

Machine Learning Studio provides multiple methods for performing feature selection.

Choose a feature selection method based on the type of data that you have, and the requirements of the statistical technique that's applied.

Feature selection

Feature are created from raw data through a process of feature engineering.

For example, a time stamp in itself might not be useful for modeling until the information is transformed into units of days, months, or categories that are relevant to the problem, such as holiday versus working day.

Learning with Counts

The **Learning with Counts** category includes the following modules:

- Build Counting Transform
- Export Count Table
- Import Count Table
- Merge Count Transform
- Modify Count Table Parameters

Related tasks

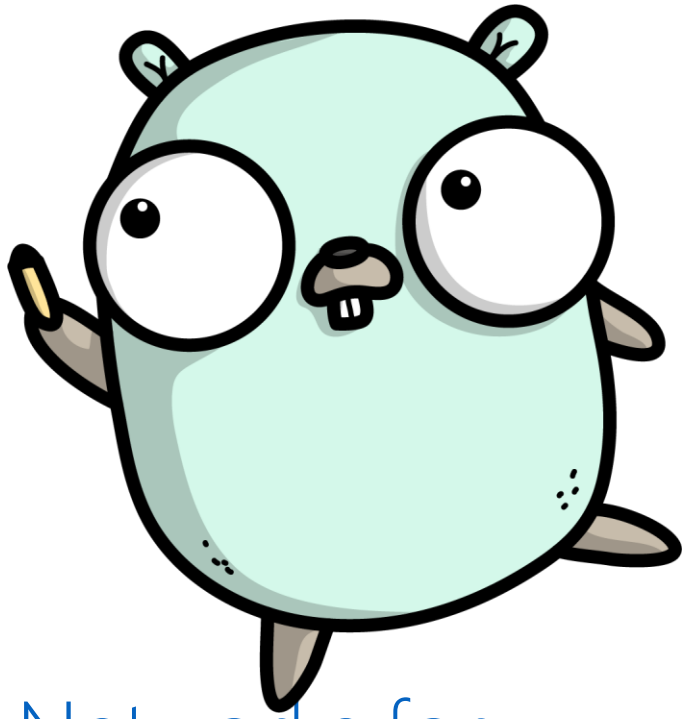
The following modules aren't in the Feature Selection category, but you can use them for related task. The modules can help you reduce the dimensionality of your data or find correlations:

- Principal Component Analysis
- Learning with Counts
- Compute Linear Correlation

Demo

Using Feature Engineering & Selection

Resources



TutorialsPoint

Microsoft Docs

Lecture Collection | Convolutional Neural Networks for Visual Recognition(Spring 2017)

Python Numpy Tutorial

Image Credits: @ashleymcnamara

Thank you

Eng Teong Cheah
Microsoft MVP Visual Studio & Development Technologies
Twitter: @walkercet
Github: https://github.com/ceteongvanness
Blog: https://ceteongvanness.wordpress.com/
Youtube: http://bit.ly/etyoutubechannel