

Feature Selection

Lesson



AI Academy

Data Preprocessing

- Sampling
- Feature subset selection
- Dimensionality Reduction
- Feature creation
- Discretization and binarization
- Attribute Transformation

Vocabulary Note

- Feature vs Attribute vs Variable
 - They all refer to “columns” in our dataset
 - “Feature” is often used when talking about removing, creating or transforming attributes.
- Class or Target Attribute
 - The attribute we want to predict

X → Y

Feature Subset Selection

- **Feature Selection:** Selecting a subset of attributes to improve the performance of a learning algorithm.
 - An important way to reduce dimensionality of data.

Feature Subset Selection

Redundant Feature

Duplicates much or all the information contained in one or more other attributes

- **Example:** Purchase price of a product and the amount of sales tax paid

Irrelevant Feature

Contains no information that is useful for the data mining task at hand

- **Example:** Student ID is often irrelevant in predicting GPA

Feature Selection Techniques

Embedded: Feature selection occurs naturally as part of the data mining algorithm

- Example: Decision Tree

Feature Selection Techniques

Filter: Features are selected before data mining algorithm is run

- Example: Correlation-based feature selection

Feature Selection Techniques

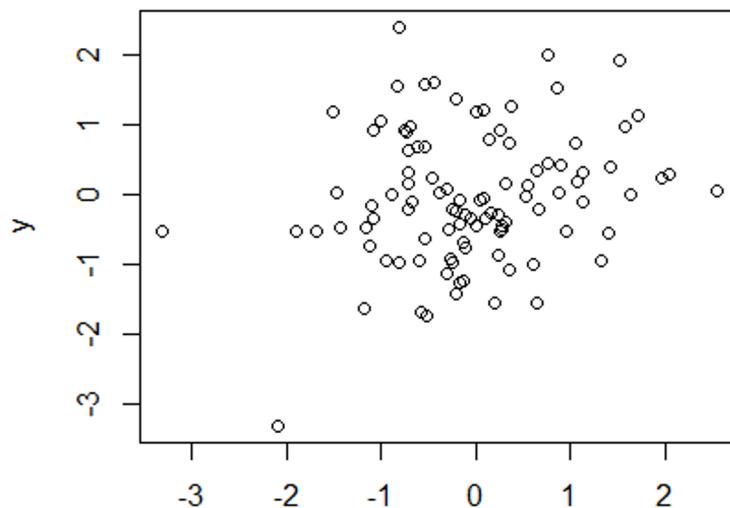
Wrapper: Use the data mining algorithm as a black box to find the best subset of attributers

- Example: Brute force feature selection

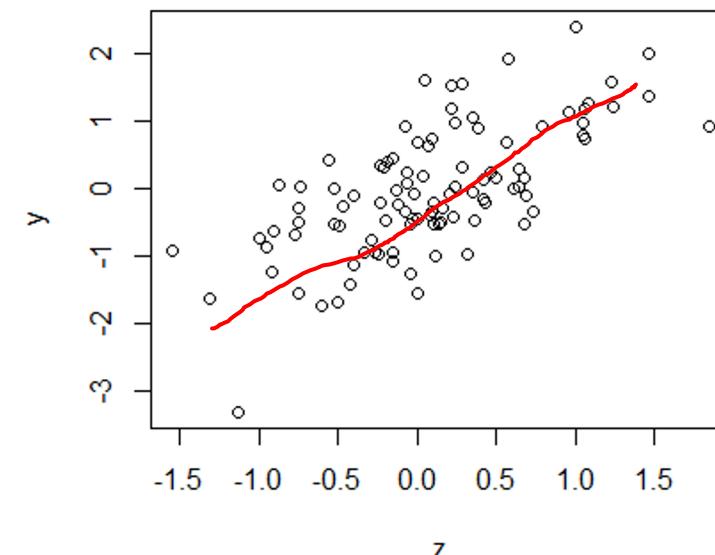
Example: Correlation Ft. Selection

Idea: Filter attributes that are not correlated with class variable

Which feature is better for predicting y?



$$\text{cor}(\text{x}, \text{y}) = \underline{0.197}$$

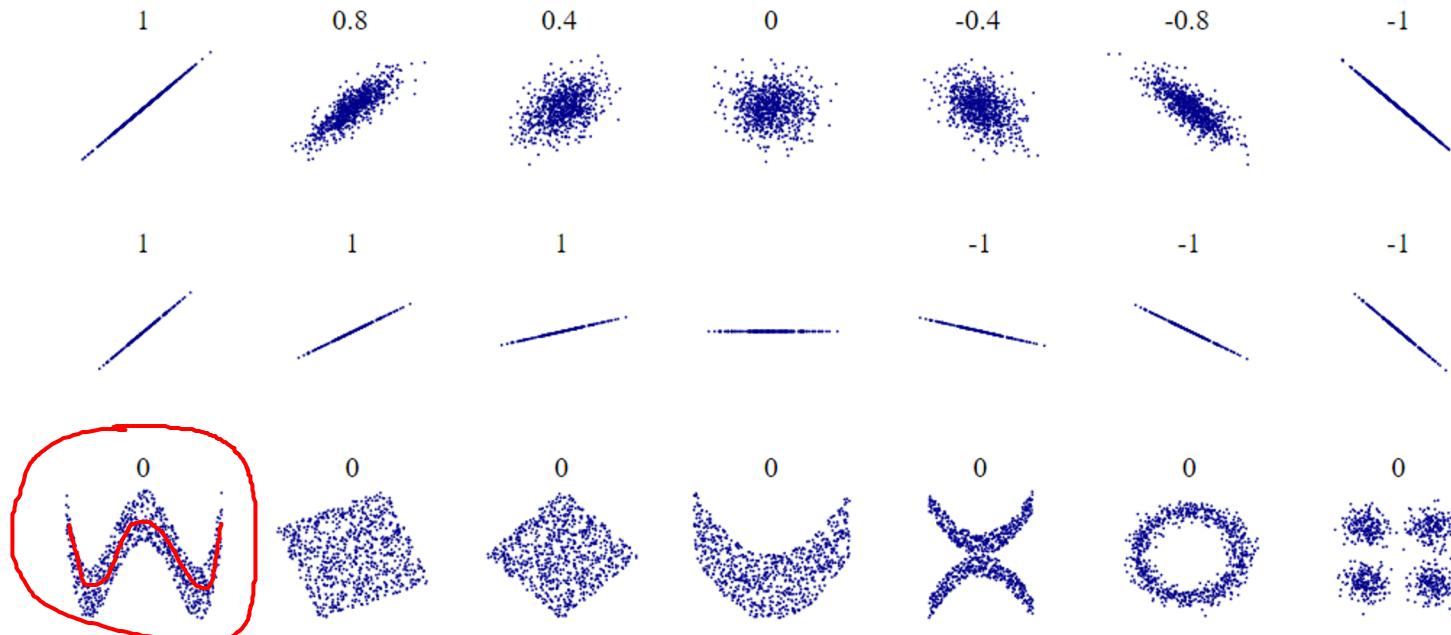


$$\text{cor}(\underline{\text{z}}, \text{y}) = \underline{0.668?}$$

Example: Correlation Ft. Selection

Warning!

- Features may be related non-linearly!
- Features may be useful only *with other features*.



Learning Objectives: Feature Selection

You now should be able to:

- Explain the motivation, advantages and challenges of feature selection.
- Apply a simple feature selection approach (e.g. correlation-based)



AI Academy
NC STATE



Feature Selection

Exercises



AI Academy



Practice: Feature Selection

(On Moodle): Which of the following is true about feature selection?