

Classification of Data Mining Problems v1

Mert Nuhoglu

July 23, 2015

Classification or class probability estimation

- ▶ What is the class of something?
- ▶ Example:
 - ▶ Is this twitter message positive or negative about your brand?
- ▶ Scoring
 - ▶ What is the likelihood that this message is positive?

Regression

- ▶ Regression: How much?
- ▶ Classification: Which class?
- ▶ Predicting future events
 - ▶ How much will the customer buy from your market?
 - ▶ Will the customer leave his subscription service?

Similarity matching

- ▶ Identify similar entities
- ▶ Example
 - ▶ Which customers are similar to current customer?
 - ▶ Purchased similar items
 - ▶ Liked similar items

Clustering

- ▶ Clustering: Group similar entities in general
- ▶ Similarity matching: similar entities to some specific entity

Co-occurrence grouping

- ▶ Co-occurrence grouping: Group entities based on the actions done unto them
- ▶ Clustering: Group entities based on their attributes
- ▶ Example
 - ▶ Market basket group: Tomato sauce and spaghetti
 - ▶ Because people buy them together
 - ▶ Although their attributes are not similar

Profiling

- ▶ Describe the behavior of people
- ▶ Example
 - ▶ Credit Card fraud detection:
 - ▶ Does this transaction match the profile of the client?

Link prediction

- ▶ Predict missing connections between entities
- ▶ Example
 - ▶ You have 10 common friends with Helen on Facebook.
 - ▶ Do you want to add Helen as friend?

Data reduction

- ▶ Explain the general data by a subset of it

Causal modeling

- ▶ Which variable causes some event?
- ▶ Example
 - ▶ Ice cream seller
 - ▶ Suspected causes:
 - ▶ Advertising
 - ▶ Season

Conclusion

- ▶ Too many different algorithms
- ▶ But only a small number of problem types