MachLe

Summary

Lukas Schöpf

23. Januar 2025

1 Introduction

1.1 Data Types

- Numerical/Quantitive, discret: Countable

  Example: Number of rejected Loans, classes taken this semster

- Numerical/Quantitive, Continuos: Interval data

  Example: Distance, mg of drug taken, size of a house

- Categorical, ordinal: distinct and can be ordered

  Example: Credit score can be {low, medium, high}

- Categorical, nominal: categories cannot be orderd

  Example: gender, eye color

1.2 Machine Learning Paradigms

- Unsupervised Learning: Discover and explore structure from unlabelled data
- Supervised Learning: Learn to predict/forecast an output of interest, we know what we want to predict and labelled data is available

1.2.1 Unsupervised Learning

Tasks:

- Dimensionality reduction
- Feature Learning
- Matrix compilation
- Anomaly detection
- Generating data

1.2.2 Supervised Learning

Given a set of features/attributes for some objects and also the ouput/target value of what we want to predict. The Supervised ML task: Given a new object and its features what would be the output value:

- Regression: Ouput is a numeric value
- Classification: Ouput is a categorical value

1.3 Data Preparation/Preprocessing

Data will rarly be in the format and quality needed for analytics and model training and several of these operations will be needed:

- Data integration/consolidation: Collects and merges data from multiple sources into coherent data store
- Data cleaning: removing or modifying incorrect data, identify and reduce noise in data
- Data transformations: normalize, discretize or aggregate the data
- Data reduction: reduce data size by reducing the number of samples or reducing the number of attributes, balance skewed data

Low quality data will result in low quality results