

Data Science Toolbox

Mike Micatka

April 13, 2016

Contents

1	Data Techniques	1
2	Machine Learning Techniques	2
2.1	Supervised Learning	2
2.1.1	Averaged One-Dependence Estimators (AODE)	2
2.1.2	Bayesian Statistics	3
2.1.3	Case-Based Reasoning	3
2.1.4	Gaussian Process Regression	3
2.1.5	Gene Expression Programming	3
2.1.6	Group Method of Data Handling (GMDH)	3
2.1.7	Inductive Logic Programming	3
2.2	Unsupervised Learning	3
2.3	Semi-Supervised Learning	3
2.4	Reinforcement Learning	3
2.5	Deep Learning	3
3	Choosing the Right Algorithm	4

What is Data Science?

This is a work in progress. Mainly made for learning purposes. Hopelessly abbreviated, will do my best to provide sources/other resources.

Chapter 1

Data Techniques

Chapter 2

Machine Learning Techniques

2.1 Supervised Learning

2.1.1 Averaged One-Dependence Estimators (AODE)

Averaged One-Dependence Estimators is a probabilistic classification technique. It is an improvement on the Naive Bayes Estimator. This technique produces class probabilities rather than single classes which allows more flexibility by having the end-user set the threshold for selection. The computational complexity for training is $O(ln^2)$ and $O(kn^2)$ for classification where n is the number of features, l is the number of training samples, and k is the number of testing samples.

Add in equation soon, too lazy to type it out now

The computational complexity makes this technique infeasible for high-dimensional data but is linear with respect to the number of samples so can handle large amounts of training data.

Implementations:

- Weka

- 2.1.2 Bayesian Statistics
- 2.1.3 Case-Based Reasoning
- 2.1.4 Gaussian Process Regression
- 2.1.5 Gene Expression Programming
- 2.1.6 Group Method of Data Handling (GMDH)
- 2.1.7 Inductive Logic Programming
- 2.2 Unsupervised Learning
- 2.3 Semi-Supervised Learning
- 2.4 Reinforcement Learning
- 2.5 Deep Learning

Chapter 3

Choosing the Right Algorithm