

Introduction to Data Mining & Machine Learning

Data Mining & Machine Learning

- ➔ Data Mining – a set of techniques for extracting knowledge from data.
- ➔ Machine Learning – a set of algorithms that can learn from data.

Machine Learning

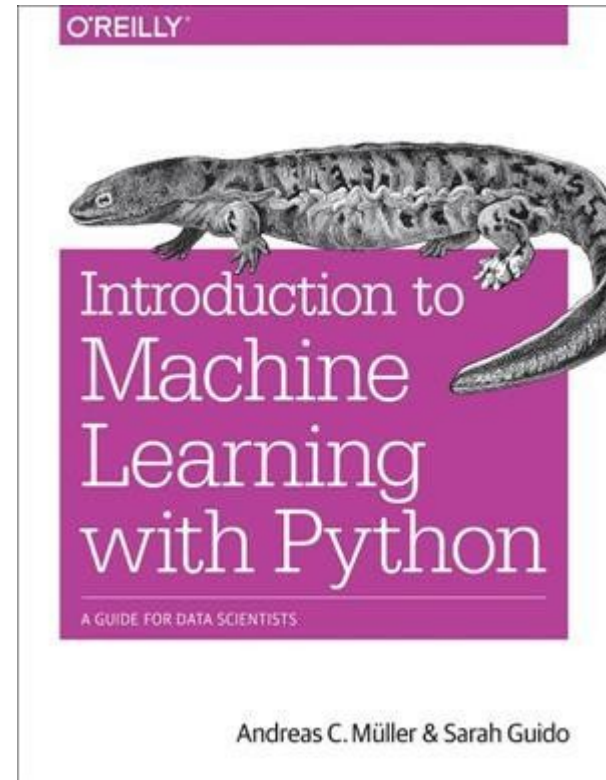
- Initially concerned concerned with robotics, computer vision, process control etc.
- Developed the algorithms.

Data Mining

- Data Mining started as an offshoot of Machine Learning.
- Tends to be more concerned with applications.
- Perhaps not as interested in the fields of robots, computer vision, process control etc.
- Uses a lot of algorithms developed as part of ML.

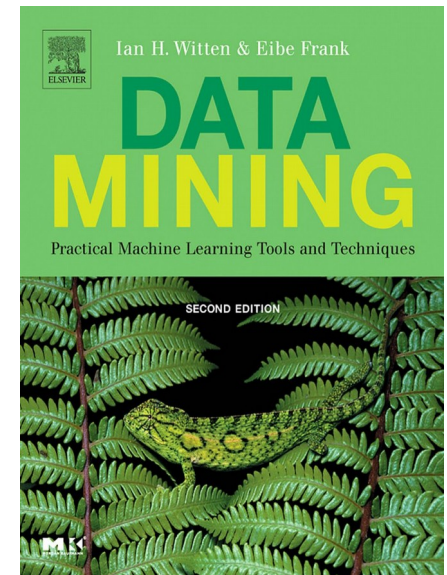
Resources

- ➔ Introduction to Machine Learning with Python : A Guide for Data Scientists
- ➔ Sarah Guido, Andreas C. Mueller



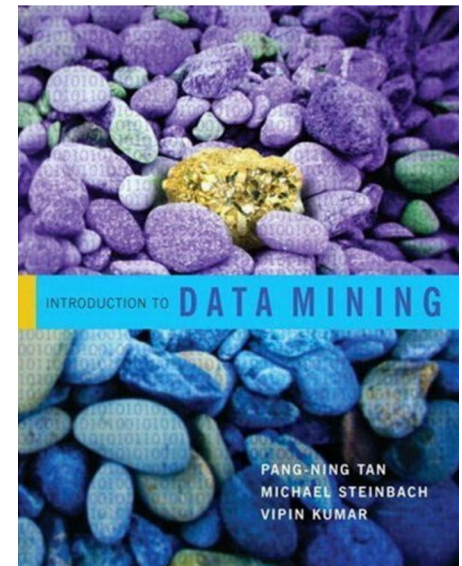
Resources

- ➔ Data Mining: Practical Machine Learning Tools and Techniques
- ➔ Ian H. Witten & Eibe Frank
- ➔ Ebook available in library.
- ➔ Slides
- ➔ Weka software
- ➔ <http://www.cs.waikato.ac.nz/ml/weka/index.html>



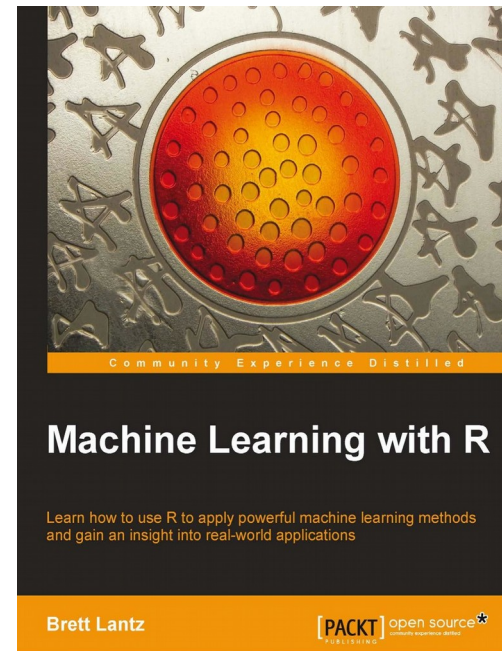
Resources

- Introduction to Data Mining
- Pang-Ning Tan & Michael Steinbach & Vipin Kumar
- Slides.
- Ebook available in library.



Resources

- Machine Learning With R
- Brett Lantz
- R Code



Lots of Opportunities

- ➔ Marketing
- ➔ Online advertising
- ➔ Recommendations
- ➔ Customer Relationship Management
- ➔ Fraud Detection
- ➔ Spam Filtering

Predicting Customer Churn

- 20% of customers leave their telecom provider when their contract expires.
- This would be an area where data mining might help predict the type of incentives that should be offered to try and hold on to customers.
- First it would need to predict the customers that are likely to leave.
- And then what incentives to offer to those customers.

Data Driven Decision Making

- Rather than relying on intuition, make decisions based on data analysis.
- Studies have shown that companies basing decision on data rather than intuition are substantially more successful.

Automated Decisions

- Sometimes the human is removed from the decision making process.
- For example,
 - fraud detection
 - spam email
 - automated recommendations (Amazon, Netflix)
 - micro financial transactions

Data as a Strategic Asset

- Signet Bank believed that with the right data they could increase profits by offering customised terms on credit cards (pricing, credit limits, cash back, loyalty points etc.)
- But they hadn't enough data.
- They decided to 'purchase' data.
- Different terms were offered at random to different customers in order to collect the data required.

Data as a Strategic Asset

- Initially they lost money (the cost of the data).
- Then its credit card operations became so successful that it was spun off as Capital One.
- One of the largest and most successful Credit Card issuers in the industry.

Data as a Strategic Asset

- Conclusion
- Data is Valuable
- With suitable data, analysis can be used to improve existing business processes.