Differences between Data Mining (572) and Statistical Models (575)

Theja Tulabandhula

572

- Introduction to data mining and R
- Decision trees
- Linear classifiers and Logistic regression
- Overfitting
- Data manipulations
- Evaluating and visualizing model performance.
- Naïve Bayes classifiers
- Neural networks
- Support vector machines
- Distance based classifiers
- Clustering
- Market basket analysis

575

By intuition, we mean 'statistical and probabilistic perspectives'

- Intuition for linear regression, bias-variance tradeoff

- Intuition for ridge regression via PCA

- Intuition for knn, curse of dimensionality

- Intuition for linear classification, assumptions

- Maximum likelihood estimation

- Sampling and Expectation Maximization

- Gradient boosted decision trees

- Geometry and Support vector machines

- Intuition for spectral clustering and PCA

- Intuition for association rules

- Time series models

There is overlap…

More importantly, they are complementary

Use off the shelf packages

Diagnose results

Design your own methods

Practice

Intuition/Theory

Explore datasets

Understand which methods are better or worse

572 Learning Goals

575 Learning Goals