



Learning Goals 575

Theja Tulabandhula



Overall Goals

Intuition while choosing between methods

Confidence while considering use of methods

Cognizance of choices being made in an analytics workflow

Successfully reason about tweaking methods in a principled manner



A Note on Practice

The goal of the assignments is to get you up to speed with R and practice with a few simple datasets

The goal of the lectures is to ask 'why' for every step we take (or choice we make) in an analytics workflow

These two complement each other so the learner can get a complete view of the subject.





575: Goals for Lecture 1

1

Understand the difference between supervised and unsupervised learning

3

Able to use vector and matrix notations for data, inputs and outputs

2

Understand what is meant by 'linear' in a linear model

4

Understand the procedure behind the nearest neighbor classifier



575: Goals for Lecture 2

1

Understand the statistical assumptions being made for k-nearest neighbor and linear models

3

Understand what the best (but unrealizable) regression and classification models are

2

Understand that curse of dimensionality is the motivation for looking at other models

4

Connect the dots between linear model and k-nn to other 'richer' models



575: Goals for Lecture 3

1

Know how to compare models using bias and variance

3

Understand the assumptions behind statistical properties of linear regression coefficients

2

How to change the bias and variance of a model family via regularization

4

Completely understand the mechanics of cross validation



575: Goals for Lecture 4

1

Why is subset selection, ridge or LASSO needed versus ordinary least squares?

3

How are they related to a likelihood interpretation?

2

What is the basic intuition behind Ridge regression? LASSO?



575: Goals for Lecture 5

1

How can you do classification via regression?

3

What assumptions underlie logistic regression?

2

What assumptions underlie linear discriminant analysis?

4

Review of bias variance tradeoff, and their relation to model selection and assessment.



575: Goals for Lecture 6

1

Review of cross validation

3

The notion of Bootstrap

2

MLE as a unifying point of view for supervised and unsupervised learning

4

Review of the first 6 lectures



575: Goals for Lecture 7

1

What is a Gaussian mixture model?

3

Why do we need to sample? How can we sample using Gibbs sampling?

2

The motivation for the Expectation Maximization algorithm

4

What are generalized linear models?



575: Goals for Lecture: Data and Decisions

1

Overview of three industry problems

3

Learn the details of a marketing decision problem and how predictions help

2

What is the newsvendor problem and how does machine learning help?

4

See the application of logistic regression for ICU mortality prediction (an application in clinical decision support)



575: Goals for Lecture 8

1

Why and how are tree based methods nonlinear? How to optimize their 'parameters'?

2

What are some practical issues with tree based models?

3

What is the idea behind boosting (with adaboost as an example)?

4

Lets demystify gradient boosting methods!



575: Goals for Lecture 9

1

What is bagging? How does Random forest improve over a single tree?

2

How to interpret a learned random forest? (variable importance)

3

Walkthrough of a nonlinear model from scratch: the Multivariate adaptive regression splines model

4

The geometric idea behind support vector machines



575: Goals for Lecture 10

1

SVMs: What are support vectors?

3

Unsupervised learning from the point of view of a joint distribution

2

SVMs: Extension to nonlinear settings via Kernel trick

4

Association rules: what is the core idea?
Extending beyond market basket data.



575: Goals for Lecture 11

1

What is dissimilarity?

3

Principal components: why are they needed? How to compute them?

2

How does K-means work? We will look at it from an optimization point of view.

4

Spectral clustering: what is spectral? spectral Understand it as a spectral decomposition followed by k-means.



575: Goals for Lecture 12

1

What is different in time series datasets compared to supervised learning datasets?

3


What are the core ideas behind the ARMA model construction?

2

How to model dependencies across time?

4

Can supervised learning methods be used for this? (hint: yes!)



**Make use of the resources, reach out to
the teaching staff and have fun learning!**

