

ECE 4200...
what we learnt

Broad topics

Supervised learning

Unsupervised learning

Recommender Systems

Hidden Markov Models

Supervised Learning

- Decision trees
- Naïve Bayes, Gaussian Naïve Bayes
- Perceptron
- Logistic regression
- Linear regression
- Support vector machines
- Boosting, Bagging
- Nearest Neighbor methods
- Neural Networks

Unsupervised Learning

- Principal components analysis
- Johnson Lindenstrauss lemma
- K-means clustering, k-means algorithm
- Hierarchical clustering
 - Single, average, complete linkage clustering

Tools

- Probability
- Linear algebra
- Optimization

Principles

- Maximum likelihood principle
- Empirical Risk Minimization

Principles

Goal: minimize the test error! Do not overfit!

- Bias variance trade-offs
- Occam's razor

Decision Trees

Entropy

Information gain

Gini criterion

- simple, easy to overfit
- Great with ensemble methods (eg Boost)

Naïve Bayes

Bayes rule + **Probability basics**

Gaussian distribution, binomials, multinomials

Gaussian NB, Multinomial NB

- probabilistic, generative method

Perceptron

Linear classifiers

Perceptron update rule

- Data should be linearly separable
- Good with ensemble methods

Regression

Linear Regression

Logistic Regression

Solving least squares

Gradient descent

- Probabilistic interpretation of the optimization

SVM

Margins

Max-margin classification

Slack variables

Kernels, polynomial, and RBF kernels

- Very powerful (particularly with kernels)
- Time consuming

Instance based learning

Nearest neighbors

- awesome, but time consuming

Ensemble methods

Bagging and Boosting

Variance reduction method

Bias reduction in boosting

Decision stumps

- Improve weak learners into strong learners

Neural Networks

Structure of Neural Networks

Sigmoid transfer function

Back propagation

Vector BP, Jacobian

- State of the art for many problems
- Black box for the most part
- Very little theory known

Dimensionality reduction

PCA, JL

Computing eigenvalues, eigenvectors

Computing errors of projection

Understanding projections onto subspaces

Clustering

K-means, k-medians clustering

Initialization, and iteration of the algorithms

Collaborative Filtering

Similar users like similar items

- Find users like the user you are dealing with
- See how they rate the item
- Output a weighted average of these ratings

Matrix Factorization

Low Rank Approximation of the ratings matrix

- SVD is optimal without missing values
- Alternative minimization (EM?)
- Latent variables govern both users and items

Hidden Markov Models

Processes have temporal features

- Markov chains model such processes for which IID assumption is too bad
- Only have observational access to the process
- How to perform inference on such processes

Three Algorithms

- Forward Backward Algorithm
- Viterbi Algorithm (several other applications)
- Baum Welch Algorithm

We hope

- You got a flavor of ML algorithms of several kinds
- Feel that you have enough understanding to not be silent in a job interview that has ML

We appreciate and understand

- All the elements of uncertainty we are facing
- The hardness of the assignments and projects
- The effort required from all of you

I sincerely hope

You all forgive me for the mess we are having with
Assignment 9

Please

Do fill the class evaluation for the semester

Thank you very much ...