10-751: Probability and Data Analysis for Computer Science

Introduction and Outline

Idea for the Course

- Probability and statistical methods are becoming ubiquitous in computer science
  - Central in theory (ex.: prime testing, one-way branching function equivalence, randomized algorithms)
  - Central in AI (machine learning, planning, stochastic search, ...)
  - Most work that has any experimental work will require it
  - Large data sets are increasingly available and used
- Example: CRC book series

Idea for the Course

- Came out of discussions with Roni Rosenfeld and others...
- Many students don't want to take several statistics courses, which may not address the topics/emphasis they need anyway.
- Need to compute!

An Experimental Course

- Want input & feedback on the topics you think are most important
- Current choice already based on input from several people
- "Service course" that can be modified to meet needs of SCS students

Is This Course for You?

- For students doing work in machine learning, probably want a more in-depth treatment
- Larry Wasserman's courses 36-725 and 36-726 are excellent.
- If you're already familiar with probability...
- But pace of the course will be fast – and not every detail will be covered in lecture
- Undergrads welcome

Three Main Parts

1. Probability
2. Statistical inference, experiments, and graphics
3. Matrix methods and optimization

Part I: Probability

- Sample spaces, events, conditioning
- Random variables
- Special distributions
- Expectations and moments
- Point estimators, maximum likelihood
- Laws of large numbers
- Markov processes
- Different views of probability

Part II: Statistical inference and experiments

- Standard error and confidence intervals
- Hypothesis testing, p-values
- Sign and rank tests
- Regression
- Cross-validation, bootstrap
- Random variate generation
- Synthetic data and controlled experiments
- Graphics and visualization

Part III: Optimization and matrix methods

- Newton's methods
- Line search and gradient ascent
- Conjugate gradient
- Linear and quadratic programming
- Solving linear systems
- Eigenvalues and eigenvectors
- Matrix decomposition methods

Texts and Notes

The main book is:

Morris DeGroot and Mark Schervish, *Probability and Statistics* Third Edition, Addison-Wesley, 2002.

We'll also use Larry Wasserman's manuscript

All of Statistics. `http://www.stat.cmu.edu/~larry`

Computing

An emphasis will be placed on computing. We will make use of three computing environments/packages:

- Matlab
- R
- The Gnu Scientific Library

Assignments

Assignments will be given every 1-2 weeks, containing a mix of pen-and-paper and computing.

All work is expected to be your own, unless any collaboration is explicitly noted. If you get help from others, you need to acknowledge it.

Course Requirements

Written assignments	30%
Exam 1 (in class)	20%
Exam 2 (in class)	20%
Exam 3 (take home)	30%

Questions?

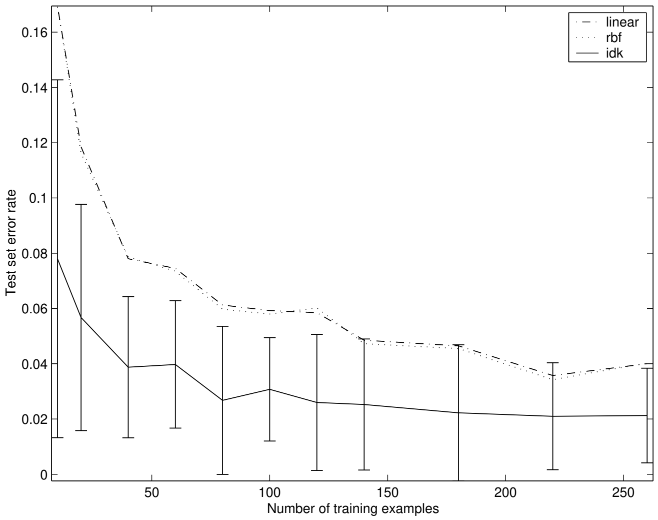
Example: Significance Tests on Experimental Results

Problem: We have a new method of ranking documents for information retrieval. Is it really better than the old method?

Significance Tests on Experimental Results

Database	*Query*	*Two-Stage*		
		AvgPr	Initpr	Pr@5d
AP88-89	SK	0.222*	0.611	0.375
	LK	0.374	0.754	**0.533**
	SV	0.204	0.598	0.368
	LV	0.292	0.689	0.473
WSJ87-92	SK	0.218*	**0.662**	**0.450**
	LK	**0.358**	**0.850***	0.607
	SV	**0.199**	**0.660**	**0.425**
	LV	**0.288***	0.762	0.520
ZF1-2	SK	0.200	0.494	0.299
	LK	0.322	**0.696**	0.434
	SV	0.181	**0.487**	**0.271***
	LV	**0.279***	**0.618**	**0.384**

Significance Tests on Experimental Results



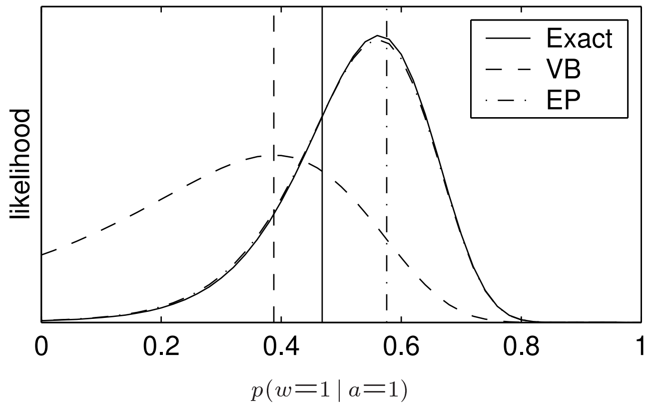
Example: Synthetic Data and Visualization

Problem: for a certain kind of mixture language model, compare two different "approximation" algorithms for estimating the parameters of the model.

Working with "synthetic data" and making visualizations is a very important tool for better understanding.

See `http://www.stat.cmu.edu/~minka/papers/aspect.html`

Synthetic Data and Visualization (cont)



Two words, two aspects, $p(w{=}2 \mid a{=}1) = 1$, $\alpha_1 = \alpha_2 = 1$

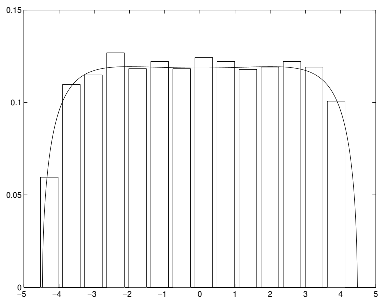
Synthetic Data and Visualization (cont)

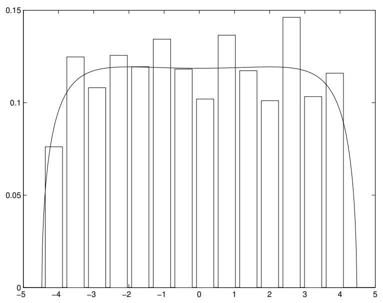
[Figure omitted]

Example: Distributions and Eigenvalues

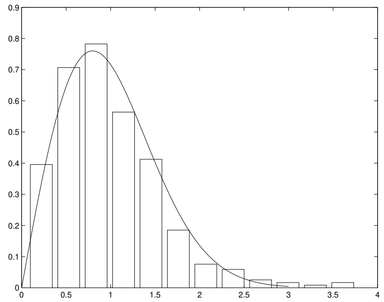
Problem: There are two types of distributions in Random Matrix Theory — Poisson and Gaussian Ensemble. Study how these arise for certain families of group representations. (Connections to the Riemann Hypothesis)

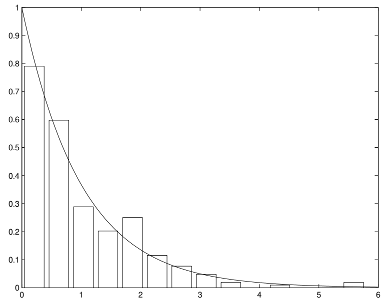
Distributions and Eigenvalues





Distributions and Eigenvalues



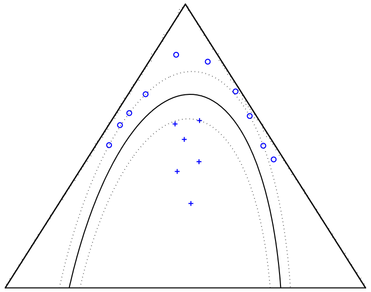


http://www.cs.cmu.edu/~lafferty/ps/cats.ps

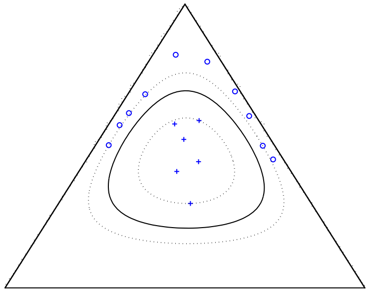
Example: Decision Boundaries for SVMs

Problem: Understand the theoretical and practical behavior of a new class of "kernels" for classification and regression.

Decision Boundaries for SVMs – multinomial

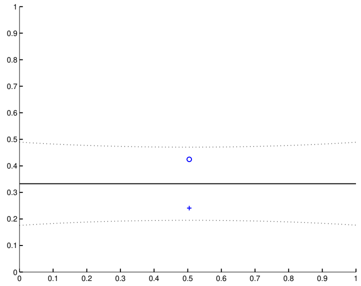


Gaussian kernel

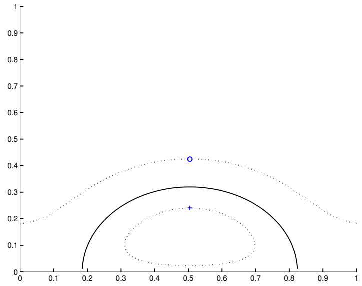


Information diffusion kernel

Decision Boundaries – spherical normal



Gaussian kernel



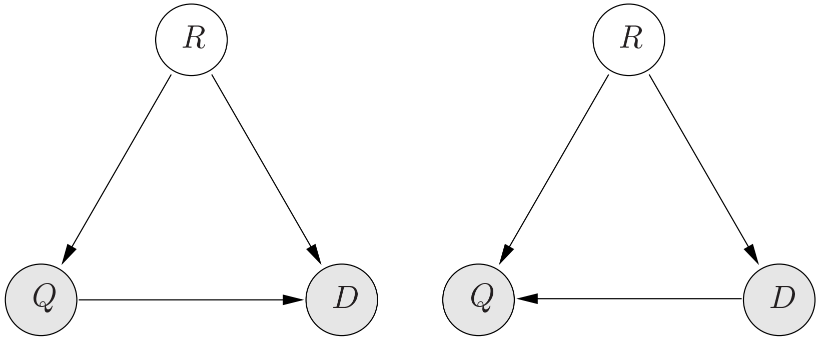
Information diffusion kernel

http://www.cs.cmu.edu/~lafferty/ps/idk.ps

Example: Equivalence of Models

Problem: We have two seemingly different probabilistic models for information retrieval. What's the relationship between them?

Equivalence of Models



http://www.cs.cmu.edu/~lafferty/ps/dq.ps

Summary

- Course thesis is that such things arise frequently in a wide range of CS research, even outside machine learning.
- Course aims to give a grounding for understanding and applying basic probability and data analysis techniques to a wide range of problems.