CSE 291 Syllabus

Spring 2017

Lecture: Monday, Wednesday, Friday 3:00-3:50 CENTR 214
Course Webpage: `http://cseweb.ucsd.edu/~dakane/CSE291/`

Professor: Daniel Kane
Email: dakane "at" ucsd.edu
Office Hours: CSE 4212 Thursdays from 2-4pm, though this will likely be flooded by CSE 101 students. If you show up around 4pm, I can stay after to talk to you. Alternatively, if this time doesn't work, send me an email and we can work out an alternative time.

Course Description: CSE 291 will focus on covering recent results in computational statistics and machine learning, focusing on problems relating to the learning and testing of distributions.

Prerequisites: I will try to make this course accessible to anyone with a solid background in linear algebra (including vector spaces and subspaces, inner products, eigenvalues, positive definite matrices and the spectral theorem for self-adjoint operators), probability theory (including probability distributions, Poisson and Gaussian distributions, expectation, variance, Chebyshev and Chernoff bounds) and algorithms. That said, the course will require a fairly high level of mathematical sophistication and will move fairly quickly.

Textbook: There will be no textbook for this course, but relevant readings will be linked to from the course webpage.

Grades:
Scribe Notes: All students taking the course for a grade will be asked to produce scribe notes for one lecture (I will try to arrange a schedule for this early in the quarter). These should be clean LaTeXnotes of the material covered in class that day.
Homeworks: Students taking the class for 4 units will also be asked to complete a couple of homework assignments. These will likely be short.
Reading Project: Students taking the class for 4 units will also be asked to complete a reading project in which they read some recent research paper in the area and write a short report on it. Students should arrange a topic with me by May 1st, and should submit a first draft by June 2nd.

Schedule: Below is a rough schedule for topics covered in the class (some may be skipped depending on timing):

Preliminaries: Basic setup for distribution learning and testing, basic information theory.
Unstructured distributions: Learning complexity, L^2 testing algorithm, identity testing, closeness testing, instance optimal testing, independence testing, robust testing, lower bounds.
One dimensional families: VC-dimension, A_k-distance, effective dimension of standard families, testing algorithms for A_k-metric, covers and metric entropy, Fourier learning algorithm.
High dimensional families: Product distributions, Gaussian distributions, graphical models, mixtures of Gaussians and products, robust statistics, SQ-algorithms and lower bounds.