

UNIVERSITY OF WATERLOO

STAT 946

Winter 2009

Kernels and Ensembles

Mu Zhu, PhD
Associate Professor

Description The support vector machine and the AdaBoost algorithm have spawned a wave of research in statistical machine learning. This course will focus on the basic ideas behind these algorithms. The first one is: we can transform many classical linear algorithms into highly flexible nonlinear algorithms by using kernel functions. The second one is: we can make accurate predictions by building an ensemble of relatively simple-minded models rather than carefully fine-tuning a single model. I will begin the course, however, with a few basic nonparametric techniques. This is partly because some students may not have any background in this field, but, even for sophisticated students of machine learning, the point that one must always control the flexibility (or complexity) of one's model in order to balance the fundamental trade-off between bias and variance still needs hammering home.

Success Story Alexandra Laflamme-Sanders took this course in 2007, and did her MMath research afterwards under my supervision. She did important and original work. Together, we published a research paper in *Neural Networks*. She now works for Sun Life Financial.

Time & Location Tu & Th, 10:00 – 11:20; MC 6027.

Requirements There will be one assignment, consisting of a number of basic exercises. After I finish my lectures, each student will be required to read and present a research paper related to the theme of the course. At the end of the term, each student will be required to submit a *very short* critical essay (limited to a maximum of 2 pages). Students will also be required to attend and evaluate each other's presentations; failing to do so will result in a penalty. Auditors are encouraged (but not required) to present a research paper as well. To receive auditing credit, the only requirement is regular attendance, including students' presentations.

Evaluation

	Assignment (20%)	Essay (30%)	Presentation (50%)	Total (100%)
Extraordinary	20	30	50	100
Excellent	19	28	48	95
Good	18	26	46	90
Acceptable	16	22	42	80
Poor	12	18	35	65
Unacceptable	08	10	22	40
Did Not Participate	00	00	00	00

Example: Suppose Adam's presentation is "good" (46); his essay is "excellent" (28), and his assignment is "acceptable" (16). His final grade for the course will be $46 + 28 + 16 = 90$.

Prerequisites One must be a graduate student in order to take this course; there are no prerequisites otherwise. Familiarity with multiple linear regression and logistic regression is helpful.

Office Hours By appointment. Email “m3zhu at uwaterloo dot ca” using “946K&E” as the subject. Each student is *strongly* encouraged to make an appointment with me before his/her in-class presentation.

Presentation

Assignment By the end of Week 4, each student will be randomly assigned a research paper and a presentation date. Afterwards, students are permitted and, in fact, encouraged to trade their assignments (research paper, presentation date, or both) with each other. However, all trading activities must terminate and all students must make their final commitments by the end of Week 8. For papers that are very rich, students are encouraged to discuss with the instructor the possibility of giving more than one presentation.

Cross-evaluation Before each student presentation, every student in the audience will be given an evaluation sheet. These evaluation sheets will serve two purposes: they will make sure that students’ final grades in this course will not be based entirely on the instructor’s opinion alone; they will also serve as attendance records. There will be a penalty for not attending other students’ presentations.

Essay Here is my suggestion for the final critical essay: Pick a topic related to the course (e.g., a topic covered in the lectures, or a related research paper); use $\frac{1}{3}$ to $\frac{1}{2}$ of the essay to give a concise summary of the chosen topic; use the remaining $\frac{1}{2}$ to $\frac{2}{3}$ of the essay to offer some constructive criticisms and outline some new research ideas. However, if you have something impressive to say, you are by no means obliged to compose your essay in this format. My suggestion is *not* meant to be a constraint.

References

Breiman L (2001). Random forests. *Machine Learning*, **45**(1), 5 – 32.

Cristianini N, Shawe-Taylor J (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.

Hastie TJ, Tibshirani RJ, Friedman JH (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Laflamme-Sanders A, Zhu M (2008). LAGO on the unit sphere. *Neural Networks*, **21**, 1220 – 1223.

Shawe-Taylor J, Cristianini N (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Zhu M (2008). Kernels and ensembles: Perspectives on statistical learning. *The American Statistician*, **62**(2), 97 – 109.