



## BOSTON UNIVERSITY STATISTICS AND PROBABILITY SEMINAR SERIES

### Google and the Vapnik-Chervonenkis Dimension

Stuart Geman

Department of Applied Mathematics,  
Brown University

April 16, 4-5pm

Mathematics and Computer Science (MCS) Building, Room 149  
111 Cummington Street, Boston  
Tea and Cookies at 3:30pm in MCS 153

**Abstract:** Google engineers routinely train query classifiers, for ranking advertisements or search results, on more words than any human being sees or hears in a lifetime. A human being who sees a meaningfully new image every second for one-hundred years will not see as many images as Google has in its libraries, all of which are available for training object detectors and image classifiers. Yet by human standards the state-of-the-art, in computer understanding of language and computer-generated image analysis, is primitive. What explains the gap? Why can't learning theory tell us how to make machines that learn as efficiently as humans? Upper bounds on the number of training samples needed to learn a classifier as rich and competent as the human visual system can be derived using the Vapnik-Chervonenkis dimension, or the metric entropy, but these suggest that not only does Google need more examples, but all of evolution might fall short. I will make some proposals for efficient learning and offer some mathematics to support them.

---

For directions and maps, please see <http://math.bu.edu/research/statistics/statseminar.html>.