1014-62-924 **John Lafferty*** (lafferty@cs.cmu.edu), Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, and **Larry Wasserman**, Department of Statistics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. *Sparse Function Estimation in High Dimensions.*

The broad goal of machine learning is to automate the iterative process of observing data, formulating a model, and making predictions, so that the computer effectively "learns" to make more accurate predictions as more data become available. Modern data sets are, however, very high dimensional. In high dimensions, many learning problems, including the classical problem of function estimation, or regression, are notoriously difficult, due to the curse of dimensionality. We first analyze the fundamental mathematical limits of sparse function estimation in high dimensions, in order to clarify the types of assumptions that must be made in order for efficient learning to be possible. We then present a simple algorithm that provably can beat the curse of dimensionality when the underlying function is sparse. The method, called "rodeo" (regularization of derivative expectation operator), is based on the general idea of thresholding the derivatives of an estimator to isolate irrelevant variables. Under certain assumptions on the regression function and sampling density, the rodeo applied to local linear smoothing can be shown to achieve the optimal minimax rate of convergence, up to logarithmic factors, as if the true relevant variables were known in advance. (Received September 26, 2005)