STATISTICS COLLOQUIUM

MONDAY, OCTOBER 31, 2011
TALK: 4:00 PM — SCIENCE CENTER RM. 309
RECEPTION: 5:15 PM — SCIENCE CENTER, 7TH FLOOR

"CLASSIFICATION OF HIGH DIMENSIONAL DATA, FUNDAMENTAL STATISTICAL METHODS"

Emanuel Parzen
Statistics Texas A&M University

ABSTRACT

Widely applicable Statistical and Machine Learning Classification problem observes variables (Y,X), Y class 0 or 1, X p-dimensional features. We start with report by Fix and Hodges (1951). We outline a new approach based on our research on fundamental statistical methods that extend from simple data to complex data and unify discrete and continuous random variables. Our approach estimates Pr[Y=1|X] by density estimation of comparison densities d(u), 0<u<1; typical d(u)=f(Q(u))/g(Q(u)), where Q(u) quantile function of G(x). We estimate f starting with g by weighted distribution formula f(x)=g(x) d(G(x)). Define comparison probability ComPr[B|A]=Pr[B|A]/Pr[B]. Bayes rule can be stated ComPr[A|B]=ComPr[B|A]. Then ComPr[Y=1|X=x]=ComPr[X=x|Y=1]=d(u), letting x=Q(u), the quantile function of X. Interpret d(u) as probability density of mid-ranks mid-distribution transform W=Fmid(X)=(Rank(X)-.5)/n; Fmid(x)=F(x)-.5p(x), p(x)=Pr[X=x]. We discuss perfect classification Pr[Y=1|X} equals 0 or 1. We discuss many approaches to estimation of d(u); emphasize MaxEnt exponential model with sufficient statistics W_j that are nonparametric statistics, means of score functions S(X) constructed for each variable X from Fmid(X). We compute Wilcoxon type W_j from CORR(Y=1,S(X)). Our modeling approach illustrates the profound chicken and egg puzzle: which comes first, the parameters or the sufficient statistics? It depends on whether parameters are scientific or statistical. We first assume independent features; extension to dependence is via nonparametric estimation of copula density functions. As example we discuss famous Hepatitus data. Our constructed score functions S(X) can also be used to perform logistic regression estimation of Pr[Y=1|X].

Joint Research Subhadeep Mukhopadhyay