

R. CLIFTON BAILEY
STATISTICS SEMINAR SERIES 2016-2017

Local Nearest Neighbour Classification with Applications to Semi-supervised Learning

Timothy Cannings

Department of Data Sciences and Operations,
Marshall School of Business
University of Southern California

Johnson Center 326 – Meeting Room B
4000 University Drive, Fairfax, VA 22030
September 29, 2017
Time: 11:00 A.M. - 12:00 Noon

Abstract:

We derive a new asymptotic expansion for the global excess risk of a local k -nearest neighbour classifier, where the choice of k may depend upon the test point. We prove that, provided the d -dimensional marginal distribution of the features has a finite ρ th moment for some $\rho > 4$ (as well as other regularity conditions), a local choice of k can yield a rate of convergence of the excess risk of $O(n^{-4/(d+4)})$, where n is the sample size, whereas for the standard k -nearest neighbour classifier, our theory would require $d \geq 5$ and $\rho > 4d/(d-4)$ finite moments to achieve this rate. Our results motivate a new k -nearest neighbour classifier for semi-supervised learning problems, where the unlabelled data are used to obtain an estimate of the marginal feature density, and fewer neighbours are used for classification when this density estimate is small. The

R. CLIFTON BAILEY
STATISTICS SEMINAR SERIES 2016-2017

potential improvements over the standard k -nearest neighbour classifier are illustrated both through our theory and via a simulation study.