

利用高維度自變量資料之存活預測

Survival Prediction Using High-dimensional Predictors

程毅豪

中央研究院統計科學研究所

Abstract

利用高維度自變量資料如基因表達(gene expression)等基因體資料進行病人存活結果變量之預測(survival prediction)是當前重要的醫學及統計研究課題，其目的在希望能依據不同病人遺傳特質的不同建立其最合適之用藥及醫療方式。此高維度資料分析對統計實務也產生了重大的挑戰，因為多數傳統統計方法不適用於自變量個數多於樣本數的情形。目前熱門的高維度自變量資料存活預測方法有 Lasso 與脊迴歸(ridge regression)，其分別利用根據 L1 及 L2 距離所進行之懲罰迴歸將迴歸係數往"0"收縮。此外，結合單變量分析之 compound covariate prediction 方法亦廣為使用。本演講另將介紹一個 compound covariate prediction 的推廣，其將多變量迴歸係數往單變量迴歸估計收縮，因此同時利用了單變量與多變量分析的訊息。本演講將經由模擬研究比較這些方法的表現，並將應用這些方法分析一組乳癌病人 (van de Vijver at al., 2002, New England Journal of Medicine) 及一組肺癌病人(Chen at al., 2007, New England Journal of Medicine)的基因表現及存活資料。合作研究者: 江村剛志(中央大學)、陳璿宇(中研院)。