Abstract:

Many practical problems are related to prediction, where the main interest is at subject (e.g., personalized medicine) or (small) sub-population (e.g., small community) level. In such cases, it is possible to make substantial gains in prediction accuracy by identifying a class that a new subject belongs to. This way, the new subject is potentially associated with a random effect corresponding to the same class in the training data, so that method of mixed model prediction can be used to make the best prediction. We propose a new method, called classified mixed model prediction (CMMP), to achieve this goal. We develop CMMP for both prediction of mixed effects and prediction of future observations, and consider different scenarios where there may or may not be a “match” of the new subject among the training-data subjects. Theoretical and empirical studies are carried out to study the properties of CMMP and its comparison with existing methods. In particular, we show that, even if the actual match does not exist between the class of the new observations and those of the training data, CMMP still helps in improving prediction accuracy. Some examples will be presented including making predictions from breast cancer genomic data samples. Additionally, some delineation of the extension to the unknown grouping structure problem will be provided. This is joint work with Jiming Jiang of UC-Davis, Jie Fan of the University of Miami and Thuan Nguyen of Oregon Health and Science University.