METRIC APPROACH BASED CLASSIFIER

Nessonova M.N.
National University of Pharmacy, Kharkiv, Ukraine
saddy_me@mail.ru

The main objective of supervised classification problems is to construct the rules (classification algorithms, or classifiers) to assign objects to one or several prescribed groups. In medical applications such kind of tasks appears when one needs to predict lethality of a disease, or to estimate patient's grade of severity, in differential diagnostics of sicknesses, gene structures analysis, image recognition, etc. There exists the category of classification methods, which use information about objects' similarity or diversity computed as based on distances between them (metric approach). As a rule these methods are in use for unsupervised classification, i.e. to solve so-called clustering or taxonomy problems. When solving a problem of metric classifier construction, classes are always treated as some subsets of the space (set) of objects. Values of feature variables serve as an object's coordinates in the space; and a classifier can be constructed as a separating surface in this space. Another grounding to build a classifier is to use distances between the object and centers of classes, or nearest neighbors in each class (and the like), to evaluate estimated probabilities of object's belonging to each class.

In contrast to the standard method described above we do not treat classes as subsets of the space of objects, but as additional qualitative feature variable which describes the object. This gives us ability to build the mathematical model for dependent variable ("class") prediction by the rest of feature variables values. The model's construction results from sequential answers for the series of questions: 1) how strong is the influence on the classification of each of the predictor variables; 2) how the weights (contributions, loadings) of predictors change in different classes; 3) how the weight of each predictor changes depending on its values.

As a result we get the classification algorithm model, based on evaluation of estimators of object's belonging to classes, which allows to comply with the influence of both quantitative and qualitative feature variables; as well as to take into account non-linearity and non-monotonicity of estimator's change depending on the change of predictor variables.

Appliance of the model developed to the number of real medical experiment originated data sets has resulted in obtaining classifiers, which have form 80 up to 95% trueness of recognition. This fact gives reasons to recommend the model suggested for further practical use and theoretical study.