

Biometrics :

Nonparametric variable importance assessment using machine learning techniques

In a regression, we want to determine which variables are most important in predicting the response. For some methods, the variable importance is naturally defined. For example, in the linear regression, standardized coefficient or the change in R coefficient are possible ways. Also, in random forest, the variable importance is naturally defined. However, these are generally method specific and cannot be generalized. In this paper, they discuss a variable importance measure that can be used with any regression technique.

A novel statistical method for modeling covariate effects in bisulfite sequencing derived measures of DNA methylation

DNA methylation describes addition of methyl group to DNA molecule which can result in change of the DNA activity without altering the sequence. Bisulfite sequencing allows the sequencing of methylation but modeling and analyzing this sequencing data is difficult. (similar to other sequencing data). This paper constructs regression based hierarchical models with specialized EM algorithm to model such data.

Evaluation of longitudinal surrogate markers

Surrogate makers are examined to determine effectiveness of a treatment. Instead of having long follow up of study participants, using surrogate marker can be more effective. Most of existing studies involve single measurement of the surrogate markers. However, in reality, surrogate marker information is normally a trajectory. In this paper, the authors propose a model-free definition for the proportion of the treatment effect on the primary outcome that is explained by the treatment effect on the longitudinal surrogate makers and methods to estimate this proportion.

JMLR :

Lasso Net : A neural network with feature sparsity

One way to increase interpretability is to use only subset of features. LASSO is used in linear regression to achieve such objective. Extending such an idea to neural network is not a simple task. In this paper, the authors introduce skip layer where feature participates in hidden layer only if the skip layer is activated. Additionally, they use modified objective function with constraints.

Empirical Bayes Matrix Factorization

Matrix factorization includes common method such as PCA and FA. Many such methods use prior assumption or constraint to induce sparsity. In this paper, the authors use Empirical Bayes. Instead of assuming prior distribution like standard Bayes method, prior is estimated from the data. This allows consideration of wide variety of the prior distribution

JSSRB :

Finite sample change point inference and identification for a high dimensional mean vector.

Consider time series or stochastic process. Change point is time where underlying probability distribution of time series or stochastic process changes. Consider mean shift model where X_i are multinormal distribution with mean and there is a change in mean at time m . Therefore, alternate hypothesis is that the change is non zero at certain m . Estimate used for testing the null hypothesis is called CUSUM. In this paper, authors introduce bootstrap method for calibrating the test.