

ON THE LIMIT IMBALANCED LOGISTIC REGRESSION BY BINARY PREDICTORS

Vincent Runge¹

StatMathAppli
September 4 to 8, 2017, Fréjus - Var (France)

In this work, we introduce a modified (rescaled) likelihood for imbalanced logistic regression. This new approach makes easier the use of exponential priors and the computation of lasso regularization path. Precisely, we study a limiting behavior for which class imbalance is artificially increased by replication of the majority class observations. If some strong overlap conditions are satisfied, the maximum likelihood estimate converges towards a finite value close to the initial one (intercept excluded) as shown by simulations with binary predictors. This solution corresponds to the extremum of a concave function that we refer to as "rescaled" likelihood. In this context, the use of exponential priors has a clear interpretation as a shift on the predictor means for the minority class. Thanks to the simple binary structure, some random designs give analytic path estimators for the lasso regularization problem. An effective approximate path algorithm by piecewise logarithmic functions based on matrix inversions is also presented. This work was motivated by its potential application to spontaneous reports databases in a pharmacovigilance context.

Keywords: path estimator, pharamacovigilance model, piecewise logarithmic approximate path, limit class imbalance, rescaled likelihood, spontaneous reports database, square exact solution.

MS classification : Primary 62J12, 62F12, 62F15; secondary 34E05, 49M29, 62P10.

<https://arxiv.org/abs/1703.08995>

¹PostDoc at Evry University. LaMME laboratory.