

MLSeq package: Machine Learning Interface to RNA-Seq Data

Gokmen Zararsiz^{1,2}, Dincer Goksuluk³, Selcuk Korkmaz³, Vahap Eldem⁴,
Izzet Parug Duru⁵, Turgay Unver⁶, Ahmet Ozturk¹

¹Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri, TURKEY

²Erciyes University, Genome and Stem Cell Center (GENKOK), Division of Bioinformatics, Kayseri, TURKEY

³Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara, TURKEY

⁴Istanbul University, Faculty of Science, Department of Biology, Istanbul, TURKEY

⁵Marmara University, Faculty of Science, Department of Physics, Istanbul, TURKEY

⁶Cankiri University, Faculty of Science, Department of Biology, Cankiri, TURKEY

¹gokmenzararsiz@erciyes.edu.tr

Abstract

MLSeq package provides several algorithms including support vector machines (SVM), bagging support vector machines (bagSVM), random forest (RF) and classification and regression trees (CART) to classify sequencing data. To achieve this, MLSeq package requires a count table, which contains the number of reads mapped to each transcript for each sample. This kind of count data can be obtained from RNA-Seq experiments, also from other sequencing experiments such as DNA or ChIP-sequencing. This vignette is presented to guide researchers how to use this package.