Statistica Sinica **33** (2023), 1-21
doi:https://doi.org/10.5705/ss.202020.0226

FEATURE-WEIGHTED ELASTIC NET: USING "FEATURES OF FEATURES" FOR "BETTER PREDICTION"

J. Kenneth Tay, Nima Aghaeepour, Trevor Hastie and Robert Tibshirani

Stanford University

Abstract: In some supervised learning settings, the practitioner might have additional information on the features used for prediction. We propose a new method that leverages this additional information for better prediction. The method, which we call the *feature-weighted elastic net* (*"fwelnet"*), uses these "features of features" to adapt the relative penalties on the feature coefficients in the elastic net penalty. In our simulations, *fwelnet* outperforms the lasso in terms of the test mean squared error, and usually gives an improvement in terms of the true positive rate or false positive rate for feature selection. We also compare this method with the group lasso and Bayesian estimation. Lastly, we apply the proposed method to the early prediction of preeclampsia, where *fwelnet* outperforms the lasso in terms of the 10-fold cross-validated area under the curve (0.84 vs. 0.80, respectively), and suggest how *fwelnet* might be used for multi-task learning.

Key words and phrases: Feature information, model selection/variable selection, , prediction.

1. Introduction

Consider the usual linear regression model: given n realizations of p predictors $\mathbf{X} = \{x_{ij}\}$, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$, the response $\mathbf{y} = (y_1, \ldots, y_n)$ is modeled as

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i,$$

with ϵ having mean zero and variance σ^2. Ordinary least squares (OLS) estimates of β_j are obtained by minimizing the residual sum of squares (RSS). There has been much work on regularized estimators that offer an advantage over OLS estimates, both in terms of prediction accuracy and interpreting the fitted model. One popular regularized estimator is the elastic net (Zou and Hastie (2005)). Letting $\beta = (\beta_1, \ldots, \beta_p)^T$, the elastic net minimizes the objective function

Corresponding author: J. Kenneth Tay, Department of Statistics, Stanford University, Stanford, CA 94305, USA. E-mail: kjytay@stanford.edu.