**BIOST 2049**

**Spring 2023**

**Research Article Dissection 3B – LASSO and Elastic Net**

**Due 2/22/23 by 11:59 via the Assignment Link posted to Canvas**

Please read the article “Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions” by Joseph O Ogutu, Torben Schulz-Streeck, and Hans-Peter Piepho found on Canvas and answer the questions below.

This is an individual assignment. You can discuss the article with your classmates, but the answers you submit should be your own.

Please be brief but write in complete sentences with appropriate interpretations (a yes or no answer will not get full credit). A few sentences is all that is necessary to answer each question.

1. What is a regularized regression model? What are the types of regularized regression models that we discussed in class that are also discussed in this paper?

Regularized regression is penalized regression. Types of regularization involve different kinds of penalties. Penalties help determine which parameter estimates to include in a model, given the data distributions and their co-occurrence in single observations. We discussed LASSO, Ridge, and elastic net. Elastic net is a mix of LASSO and Ridge penalties determined by mixing factor alpha (from 0 to 1)

1. Is there anything different about the regularization methods discussed in this paper as compared to class (only focus on the ones we discussed in class)? If so, state the differences (in words, no need for formulas).

For elastic net, they multiplied their argmin function by 1 + lambda / n

1. Briefly describe the 6 methods compared in this paper and state where they do well and where they do not (if mentioned).

Ridge is preferable to lasso when variables are highly correlated because lasso does not understand context and does not know the meaning of eliminating a variable that is highly correlated to a covariate already in the model. Elastic Net extends lasso to automate mixing lasso and ridge or using ridge completely, especially when variable correlations are high. Three other methods were Ridge regression BLUP (REML of penalty so similar stability to Ridge), adaptive LASSO (unstable in the face of high collinearity), and adaptive Elastic Net (similar to stability of elastic net, but better asymptotic properties).

1. What did this paper use as a measure of predictive accuracy? What did we use in class when discussing LASSO and ElasticNet?

We used MSE. They use RMSE.

1. In the results/discussion section, did the authors discuss situations that would lower the performance of these methods? If so, state them.

Using true breeding values instead of true genomic values lowers performance of these methods in predicting GEBVs.

1. Were you aware of these types of methods before discussed in class?

I was aware of lasso, ridge, elastic-net, and cross-validation from Machine Learning (INFSCI 2595). I was interested to learn about the adaptive versions of these regularized regression methods.