

Northeastern University

College of Professional Studies

Regularization

Overview and Rationale

In order to consolidate your theoretical knowledge into technique and skills with practical and applicational value, you will use the *glmnet()* package in R to implement LASSO function to build linear and logistic models through LASSO over values of regularization parameter lambda.

Course Outcomes

This assignment is directly linked to the following key learning outcomes from the course syllabus:

- Conduct regularization method for models to describe relationships among variables and make useful predictions

Assignment Summary

Use one of the real world example data sets from R (not previously used in the R practice assignment) or a dataset you have found, to build regularization models by using Lasso (least absolute shrinkage and selection operator) and extend Lasso model fitting to big data that cannot be loaded into memory. You will fit solution paths for linear or logistic regression models penalized by Lasso over a grid of values for the regularization parameter lambda.

Use the resources in this module to guide your R code development.

Report

Your assignment/project should have a good cover/title page, introduction of what the goals of the project and the methods you use. It also should follow APA format with at least 1000 words (excluding title page and references page) and references page. In the body of your project you should incorporate the R codes and R outputs with interpretation of your results. Be sure to show all the elements in the official hypothesis, including the null and alternative hypothesis, the critical values, calculation of the test statistics and p-values. Finally, you need to make sense of your results to make good points with proper

Northeastern University

College of Professional Studies

conclusions, to show your understanding of the course material and its application to the dataset.

Graphs, figures, charts, tables are very useful to increase visual effects to impress your readers. You also should do your best to give insight and understanding to the project with a good conclusion. Please use subtitles to make your assignment more reader friendly as well.

Northeastern University

College of Professional Studies

Assignment Rubric

Category	Meets Standards	Approaching Standards	Below Standards
Introduction	Introduction provides a brief and intelligible overview of the goals and methods of the assignment	Introduction provides an overview of the goals and methods of the assignment, but is ambiguous or not concise	Does not introduce project goals, project questions or methods.
Analysis	Provides all R code and the outputs. Includes interpretation of the output, graphs, figures, charts, and tables and the significance of the results in the analysis.	Provides R codes and outputs, but the R code does not match the outputs or is missing some code or outputs. Includes limited interpretations, charts, and tables and the significance of the results in the analysis.	Does not provide R code or its outputs or minimal R code is provided. Includes few interpretations, charts, or tables. Does not identify the significance of the results in the analysis
Data Visualizations	Data visualizations are appropriate for the level and type of analysis. Graphs, figures and tables communicate insights and significance to the reader.	Data visualization are useful for the level and type of analysis, but graphs, figures and tables do not clearly communicate significance of the results to the reader.	Data visualization are used minimally or not at all. If graphs, figures and tables are used, it is unclear what they are intended to communicate or why.
Interpretation & Conclusions	The conclusion summarizes and makes sense of the results, making good points that reflect clear understanding of the assignment material.	The conclusion summarizes and makes sense of the results, making good points that reflect a basic understanding of the assignment material.	The conclusion does not summarize or attempt to make sense of the results. Conclusions do not reflect an understanding or reflect a misunderstanding of the material
Report: Writing Mechanics, Title Page, & References	There are no noticeable errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains approximately of 1000 words	There are very few errors in grammar, spelling, and punctuation; and completely correct usage of title page, citations, and references. The report contains approximately 1000 words	There are more than five errors in grammar, spelling, and punctuation; or the usage of title page, citations, and references are incomplete; or the report contains far less than 1000 words

Northeastern University

College *of* Professional Studies