

Assignment 7

Due Wednesday by 11:59pm **Points** 2 **Submitting** a file upload

Available Feb 20 at 12am - Mar 1 at 11:59pm 10 days



Assignment 7

Shortcut to [Module 7](#)

I

Do the following to create, execute and **compare** linear regression models for the price from the automotive data set. Use all features for each of your models. Make sure you remove the **symbolizing** and **normalized losses** columns first.

1. **Stepwise Regression.** Use the stepAIC function from the MASS package to perform stepwise regression starting with the linear model including all features. Which features are remain in the model? Evaluate this model using the summary and plot methods for your model object.
2. **SVD Regression.** Create a model matrix for all features using the model.matrix function with no intercept term (e.g. -1 in your model formula). Examine the first few rows of the model matrix. Notice how the categorical variables are encoded by a series of dummy variables. Does the coding make sense given the number of unique levels? Next, compute the SVD of the model matrix. Examine the singular values and determine which diagonal elements of the inverse singular value matrix should be set to zero. A plot may help you understand the fall-off in singular values. Then compute the weight vector. Use the weight vector to compute scores and evaluate the model. You may need to try models with several different numbers of inverse singular values set to zero to find a model with good overall performance.
3. **Elastic Net Regression.** Elastic net regression is the combination of ridge regression and lasso regression. In this case use an alpha parameter of 0.5 to give equal weight to each regularization method. Using the model matrix you created for part 2 and the 1-d matrix of heating loads, compute a Gaussian regression model using 20 values of the regularization parameter lambda. Compute the scores using the predict method. Plot and examine the evaluation of the parameter values and deviance with lambda. Choose a value of lambda (a column in the scores matrix) and evaluate your model. You may need to try several lambda's to find one which gives good overall performance.

To **evaluate and compare** these models use both summary statistics and plots of the residuals. How similar or different is the performance? Do you consider any of these models to be a good fit to the data?

Note, we are ignoring the obvious bias problem which arises from evaluating a model using the same data used to train the model.

You should submit:

> One R-script

- Follow good coding practice; minimize cut and paste
- Include comments

Note, if you have done your work in rmarkdown or a notebook and your code is inline with your text you do not need to submit a separate R file.

> A clear report, **stating your conclusions up-front** and supporting the conclusions with evidence in the form of charts and numeric tables.

To submit your assignment:

1. Use the "Submit Assignment" link located on the top right corner.
2. Click "Choose file" button to find and select your saved report.
3. Press the "Submit Assignment" button to turn in your assignment.