Name: Krista Miller

**Q1: Lambda in ridge and lasso regression**

1. What is the potential range of lambda?

0 to infinity

1. What is true about ridge vs lasso regression when lambda = 0?

The model is the same as OLS (least squares fit) model (there is no coefficient shrinkage when lambda = 0).

1. As lambda gets bigger, what happens to the coefficients? How does this differ between ridge and lasso?

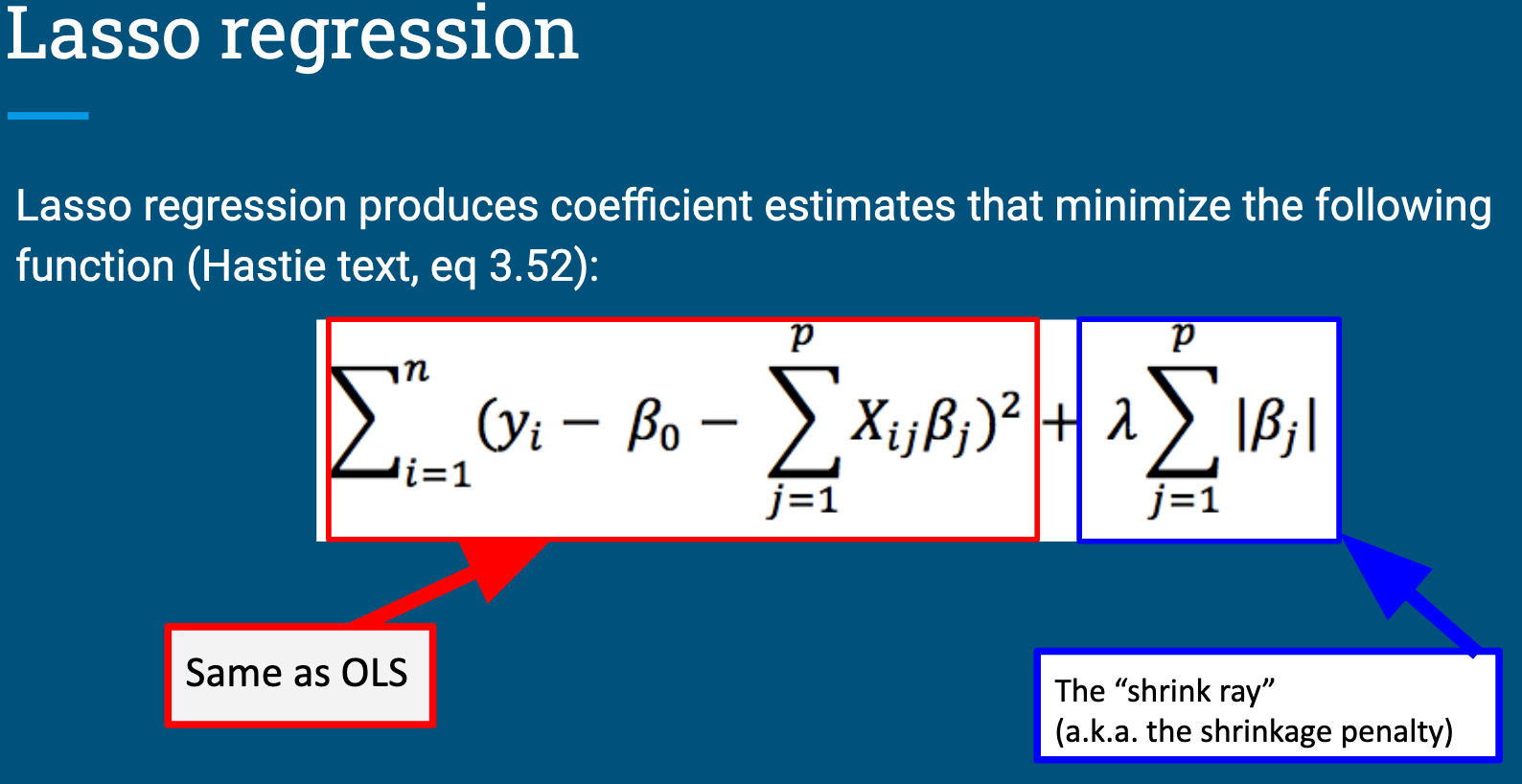
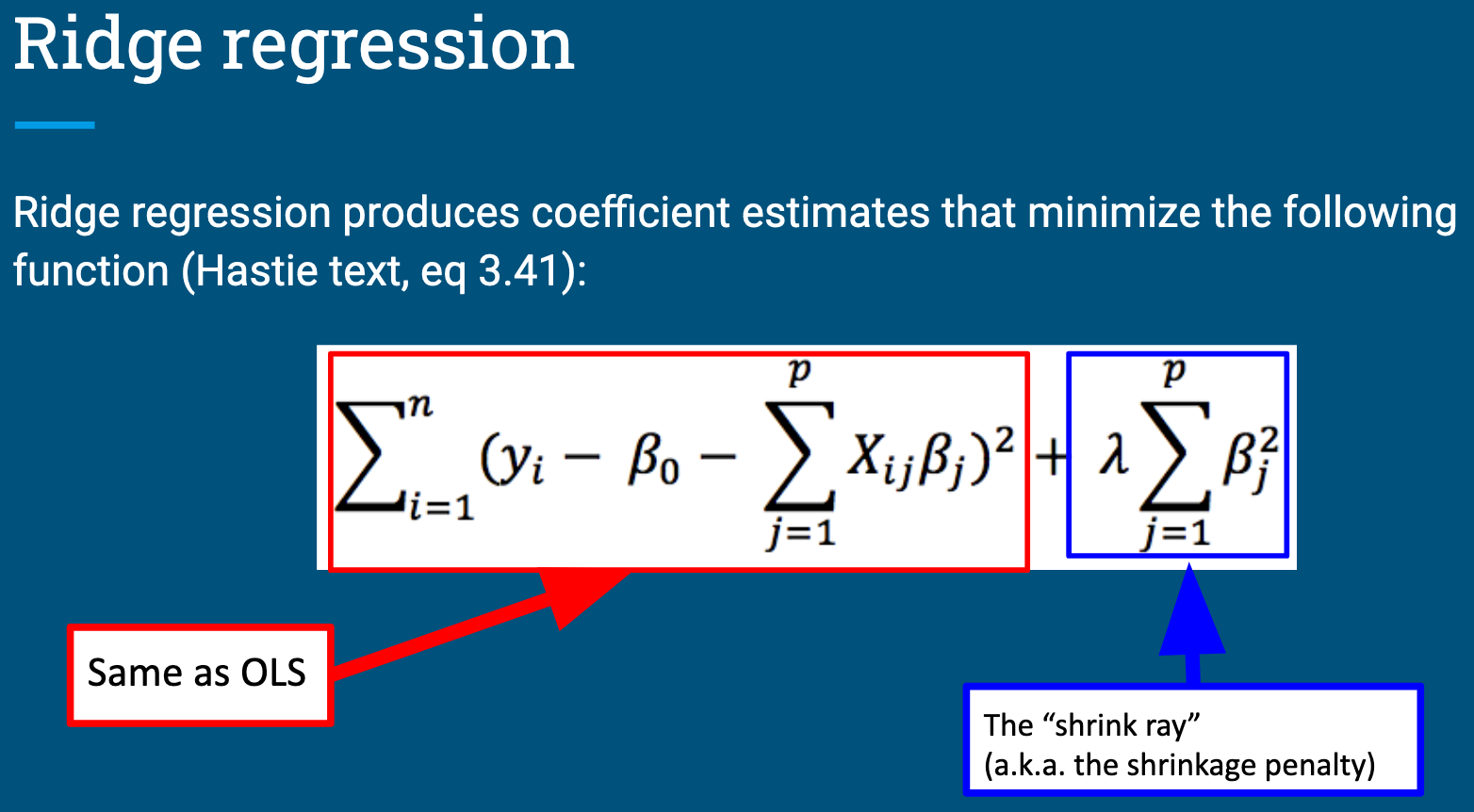
Ridge models: as lambda increases, the bias is unchanged but the variance drops. Ridge keeps all variables and shrinks the coefficients towards zero.

Lasso: as lambda increases, shrinkage occurs so that the variables that are zero can be thrown away. A major advantage of lasso is that it is a combination of both shrinkage and selection of variables.

1. What is meant by “sets of coefficients” associated with lambda?

The sets of coefficients associated with lambda are the (non-zero) weights of the predictors that are included in the final model.

Reminder of function that is minimized in each method:



There are questions on the next page – keep going!

**Q2: The glmnet() function**

The answers to the following questions can be found in the glmnet documentation. You can find this by installing the package and typing ?glmnet into the console or by searching online.

There is also a nice introduction to using the package maintained by the developers: <https://glmnet.stanford.edu/articles/glmnet.html> . This may help with understanding how to apply the demonstrations in the async to the problem set.

1. The glmnet() function in the glmnet package can fit both lasso and ridge regression models. Which one does the function perform by default? What change do you need to make if you want to perform the non-default option?

The alpha argument selects the regression. Alpha = 1 is lasso regression (default), and alpha = 0 is ridge regression.

1. The first argument in the function is ‘x’. What should you put here and what form does it need to be in?

x: matrix of predictor variables

This should be created using the function model.matrix() allowing to automatically transform qualitative variables into dummy variables, because glmnet() can only take numerical, quantitative inputs.

1. The second argument in the function is ‘y’. What should you put here and what form does it need to be in?

y: outcome variable. It needs to be quantitative for family= “gaussian” or family = “poisson”. For family = “binomial”, it should be a factor with two levels or a two-column matrix.

1. The third argument in the function is ‘family’. What is the default of this? When would you need to change it?

family = is a character string of one of the following options: “gaussian”, “binomial”, “poisson”, “multinomial”, “cox”, “mgaussian”

Gaussian is the default. You would need to change the type to binomial or multinomial if the outcome variable was in a factor format.

1. By default, how many sets of coefficients will be evaluated? What change would you need to make if you wanted to evaluate 500 lambdas?

By default, 100 lambda values will be evaluated. If you wanted to evaluate 500 lambdas, you would need to set the argument nlambda= 500

Helpful resources:

Notes:

\*ridge\_regression\_motivation\_7\_1\_1.pdf

-what does the graph on page 6 suggest about multiple regression?

multiple co-linearity is a potential problem, ridge regression/lasso is a potential solution

http://www.sthda.com/english/articles/37-model-selection-essentials-in-r/153-penalized-regression-essentials-ridge-lasso-elastic-net/