Introduction:

Linear regression is a simple method in machine learning, used to create a fit of a dataset of which it can produce continuous predictions. This is achieved by assuming that we can model the output values as a linear combination of continuous functions of the input values. For example, we can assume that the observed values in our dataset can be approximated by different polynomial degrees of the input values and find the linear coefficients $\beta$ which minimize the difference between our observed and predicted values. This general idea that we can create a model which approximates our observed data with a minimal difference between observed and predicted values is a problem at the very heart of machine learning.

However, in order to create an optimal model, we require optimal parameters which cannot be known a priori. Thus, we require extensive testing and tuning of parameters to inform our selection. This presents a major obstacle, as we must for example find the optimal polynomial degree $N$ for which our model best approximates the observed values at the same time as we tune the hyperparameter $\lambda$ responsible for an adequate regularisation. We must find these parameters which lie at an intersection, where all parameters yield the minimal error in unity.

Furthermore, we will compare different models of linear regression, namely Ordinary Least Squares (OLS), Ridge and Lasso regression, and employ resampling methods, namely bootstrap and cross-validation to ensure that we draw conclusive results as these help us reduce the randomness in our parameters based off how we split our data. <source?>

METHOD

BIG THEORY ABOUT OLS -> WHATS DIFFERENT ABOUT RIDGE AND FINNA LASSO