2️⃣

Linear Regression

Status	Done
Files	
Link	
Type	All Around
Course	Data Analysis

R Squared
Adjusted R Squared
Bias Variance Tradeoff
Regularisation
Step wise selection

R Squared

- A measure for how well the model fits the data, its better than RMSE because RMSE depends on the context
- R^2 is always between 0 and 1, and explains what proportion of the variance can be explained by the model
- The higher the better, so 1 is perfect and the model explains all variance

$$SSR = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \qquad \text{and} \qquad TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2.$$

$$R^2 = 1 - \frac{SSR}{TSS}.$$

Adjusted R Squared

- Similar to r squared but takes into account the number of predictors used, with the intuition that the simpler the model the better

$$\text{adj.}R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

- n is number of samples, and p is number of predictors

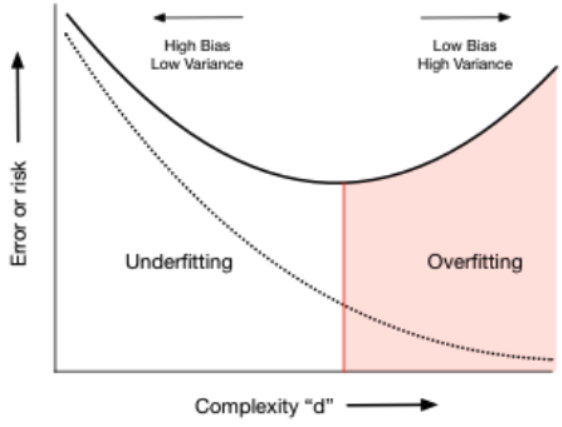
Bias Variance Tradeoff

This phenomenon is often described in terms of the *bias-variance tradeoff*. Here, we decompose the error of the model into three terms:

$$\text{Error} = \text{Bias} + \text{Variance} + \text{Irreducible Error}.$$

- The *bias* of the method is the error caused by the simplifying assumptions built into the method.
- The *variance* of the method is how much the model will change based on the sampled data.
- The *irreducible error* is error in the data itself, so no model can capture this error.

The dashed line shows us the testing error and the solid line the training error.



Regularisation

- Make the model more generalisable / less overfit to the training data, by shrinking the weight of features, testing the effect of different features using cross validation and shrinking by lambda, lambda is usually set by the user and you have to test out different options, the higher the lambda the less the weights will become as its a penalty
- Ridge $\lambda \times slope^2$ can shrink weights asymptotically close to 0 but not 0
- Lasso $\lambda \times |slope|$ can shrink weights to become 0, making the final equation simpler and easier to interpret, better when there are many variables that dont matter

Step wise selection

- Choose which features to use incrementally
  - Either starting with one and adding one by one
  - Or starting will all and removing one by one