Ridge Regression

- Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity.
- Multicollinearity occurs when there are high correlations between two or more predictor variables.
- One predictor variable can be used to predict the other.
- This creates redundant information, skewing the results in a regression model.
- Multicollinearity can adversely affect your regression results.
- It’s more common for multicollineariy to rear its ugly head in observational studies it’s less common with experimental data.
- When the condition is present, it can result in unstable and unreliable regression estimates

What Causes Multicollinearity?

- **Data-based multicollinearity:** caused by poorly designed experiments, data that is 100% observational.
- Some cases variables may be highly correlated (usually due to collecting data from purely observational studies) and there is no error on the researcher’s part.
- For this reason, you should conduct experiments whenever possible, setting the level of the predictor variables in advance.
- **Structural multicollinearity**:
- It is caused by the researcher, creating new predictor variables.

Causes for multicollinearity

- **Insufficient data**. In some cases, collecting more data can resolve the issue.
- **Dummy Variables** may be incorrectly used.
- For example, the researcher may fail to exclude one category, or add a dummy variable for every category (e.g. spring, summer, autumn, winter).
- Including **a variable in the regression that is actually a combination of two other variables**.
- For example, including “total investment income” when total investment income = income from stocks and bonds + income from savings interest.
- **Including two identical (or almost identical) variables**.
- For example, weight in pounds and weight in kilos, or investment income and savings/bond income.
- **Variance Inflation factors**

Variance Inflation factors

- A variance inflation factor(VIF) detects multicollinearity in regression analysis.
- The VIF estimates how much the variance of a regression coefficient is inflated due to multicollinearity in the model.
- VIFs are usually calculated by software, as part of regression analysis.
- VIFs are calculated by taking a predictor, and regressing it against every other predictor in the model.
- This gives you the R-squared values, which can then be plugged into the VIF formula.
- "i" is the predictor you're looking at (e.g. x_1 or x_2):

$$\mathrm{VIF} = \frac{1}{1 - R_i^2}$$

Interpreting the Variance Inflation Factor

- Variance inflation factors range from 1 upwards.
- The numerical value for VIF tells you (in decimal form) what percentage the variance (i.e. the standard error squared) is inflated for each coefficient.
- Example, a VIF of 1.9 tells you that the variance of a particular coefficient is 90% bigger than what you would expect if there was no multicollinearity — if there was no correlation with other predictors.

- A **rule of thumb** for interpreting the variance inflation factor:
- 1 = not correlated.
- Between 1 and 5 = moderately correlated.
- Greater than 5 = highly correlated.

- When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value.
- By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors.
- It is hoped that the net effect will be to give estimates that are more reliable.

- Ridge regression is an extension of linear regression where the loss function is modified to minimize the complexity of the model.
- This modification is done by adding a penalty parameter that is equivalent to the square of the magnitude of the coefficients.

- Loss function = OLS + alpha * summation (squared coefficient values)

Lasso Regression(least absolute shrinkage and selection operator)

- It is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces.
- This uses shrinkage .
- Shrinkage is where data values are shrunk towards a central point, like the mean .
- The lasso procedure encourages simple, sparse models that is model with fewer parameters.
- This particular type of regression is well-suited for models showing high levels of muticollinearity.
- Or when you want to automate certain parts of model selection, like variable selection/parameter elimination..

L1 Regularization

- Lasso regression performs L1 regularization, which adds a penalty equal to the absolute value of the magnitude of coefficients.
- This type of regularization can result in sparse models with few coefficients.
- Some coefficients can become zero and eliminated from the model. Larger penalties result in coefficient values closer to zero, which is the ideal for producing simpler models.
- Whereas the other hand, L2 regularization (Ridgeregression) *doesn’t* result in elimination of coefficients or sparse models.
- This makes the Lasso far easier to interpret than the Ridge.

Performing the regression

- The goal of the algorithm is to minimize the sum of squares with the constraint $\Sigma\ |B_j \leq s$.
- Some of the βs are shrunk to exactly zero, resulting in a regression model that's easier to interpret.
- The tuning parameter λ controls the strength of the L1 penalty.
- λ is basically the amount of shrinkage:
- When $\lambda = 0$, no parameters are eliminated.
- As λ increases, more and more coefficients are set to zero and eliminated).