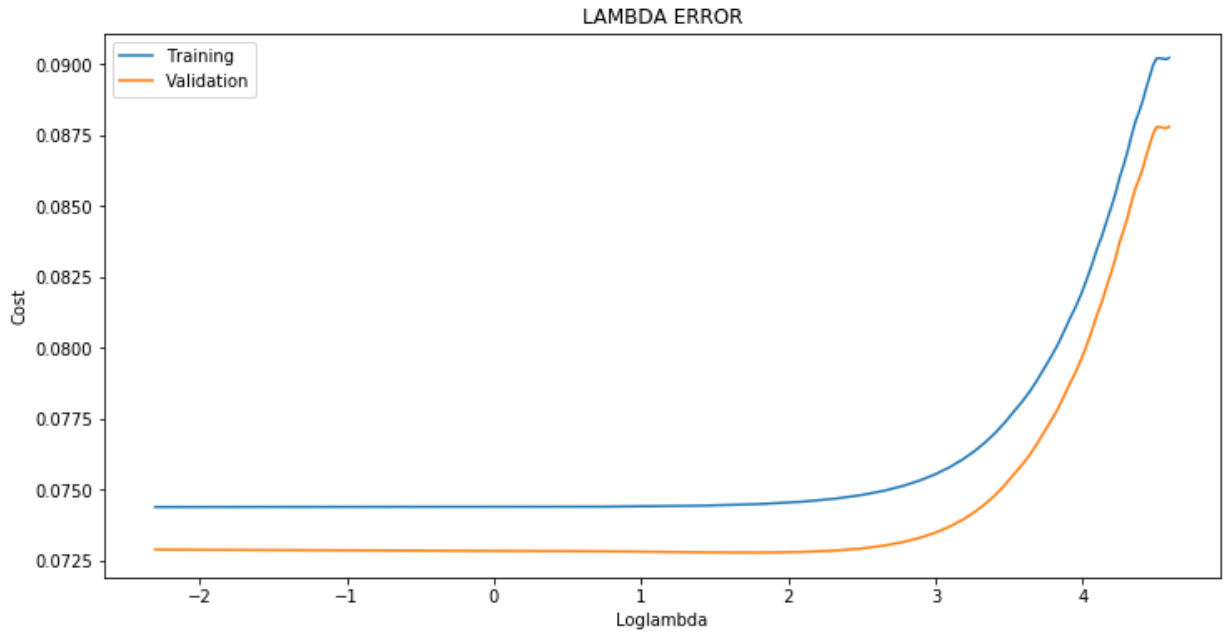
QUESTION 1
Implement L1 Regularisation (Lasso). Plot graph between regularisation coefficient λ and error

- The formula used to update weights is:

$$\theta j := \theta j - \alpha . \frac{1}{m} \Sigma (h\theta (x(i)) - y(i)) x(i) j - \lambda . sgn(w_j) . \frac{1}{2 * m}$$

Lasso selects the only some features and reduces the coefficients of others to zero.



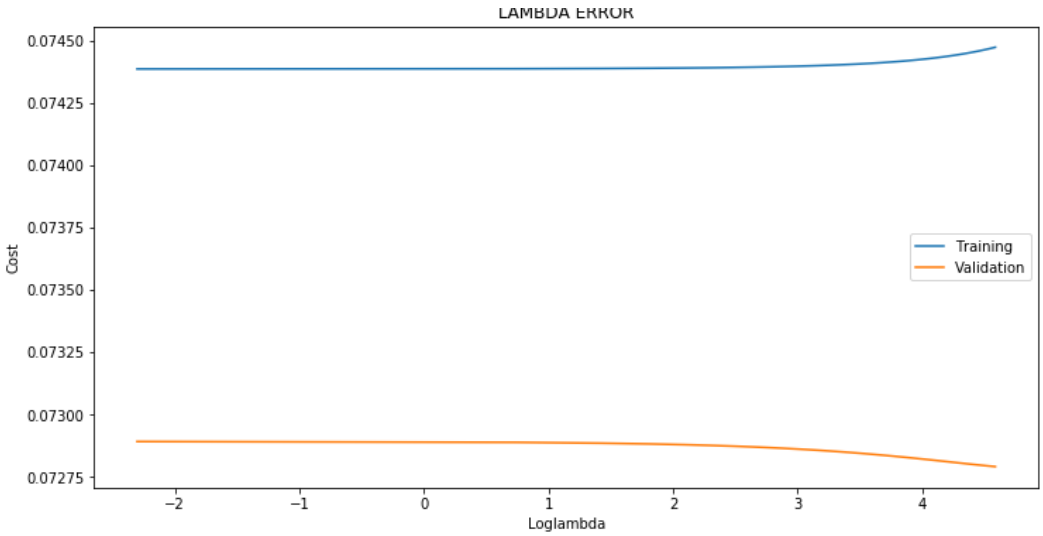
α = 0.001 λ = (0.1 to 100 in steps of 2)

QUESTION 2
Implement L2 Regularisation (Ridge). Plot graph between regularisation coefficient λ and error

- The formula used to update weights is:

$$\theta j := \theta j - \alpha . \frac{\lambda}{m} \Sigma (h\theta (x(i)) - y(i)) x(i) j - \alpha * \lambda * \frac{w_j}{m}$$

It reduces the model complexity by coefficient shrinkage.



α = 0.001 λ = (0.1 to 100 in steps of 2)

QUESTION 3

Analyse how the hyper-parameter λ plays a role in deciding between bias and variance.

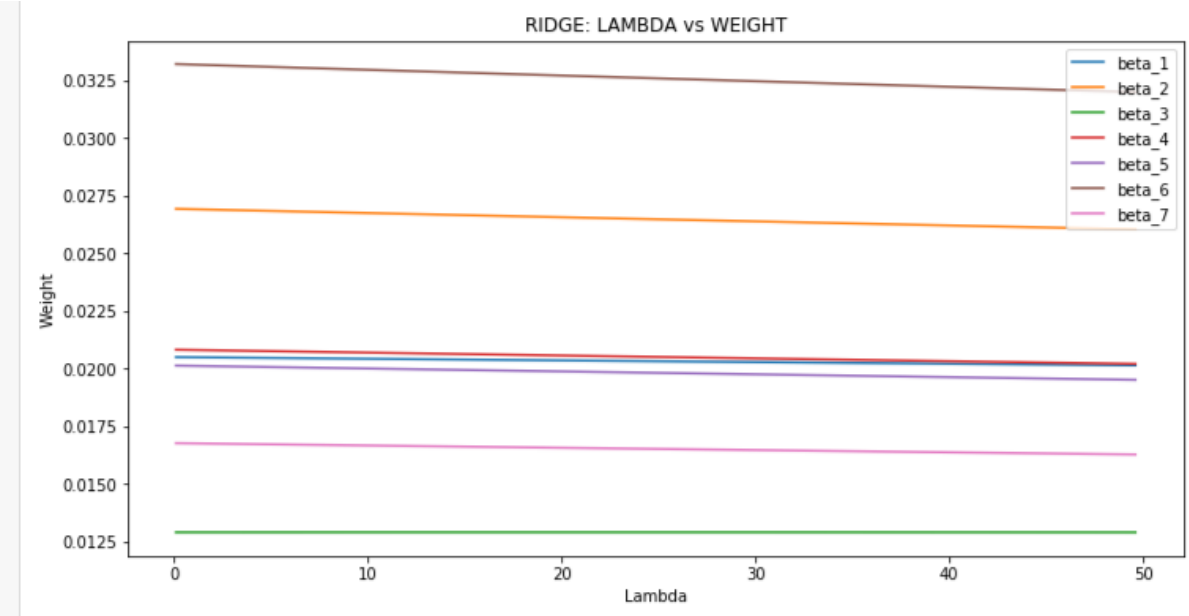
- A higher value of λ means smaller coefficients of the model. This results in much simpler model
- When λ reaches values close to infinity, underfitting takes place.
- This means bias increases with increase in λ.
- With increase in λ model generalizes in a better way, hence decreasing variance.

QUESTION 4

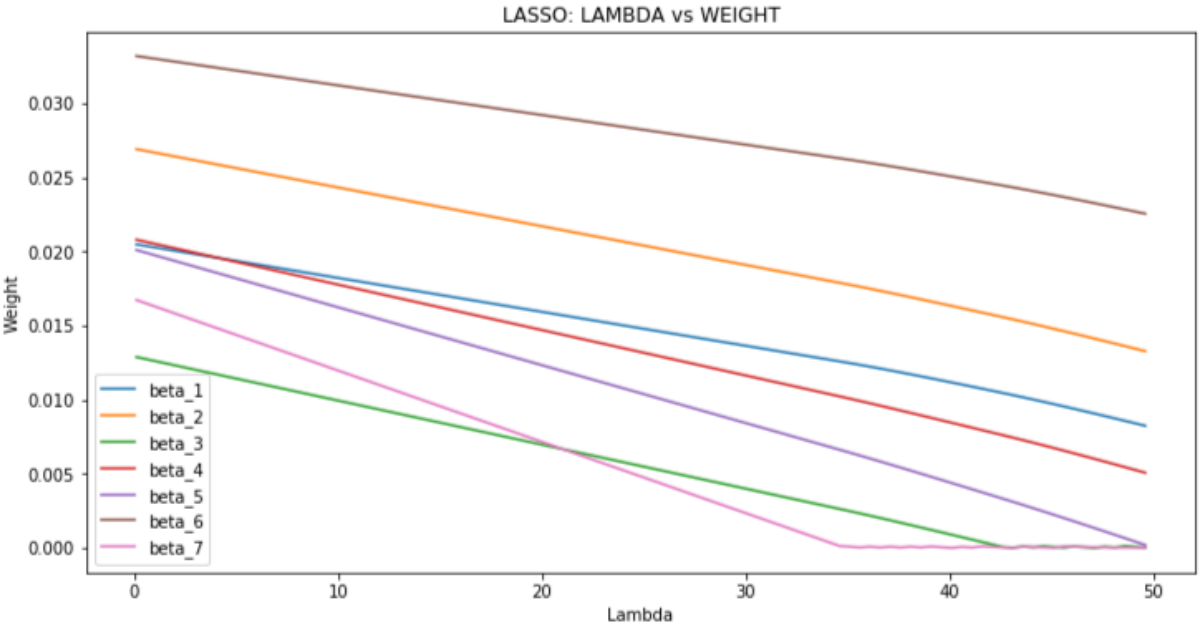
Analyse how the two different regularisation techniques affect regression weights in terms of their values and what are the differences between the two.

Using α = 0.001, number of iterations = 1000 and λ in the range of 0.1 to 50 with steps of 0.5.
- We plot values of different beta coefficients with the corresponding λ.

- In case of Ridge Regression, weights appear to vary linearly with little variation.



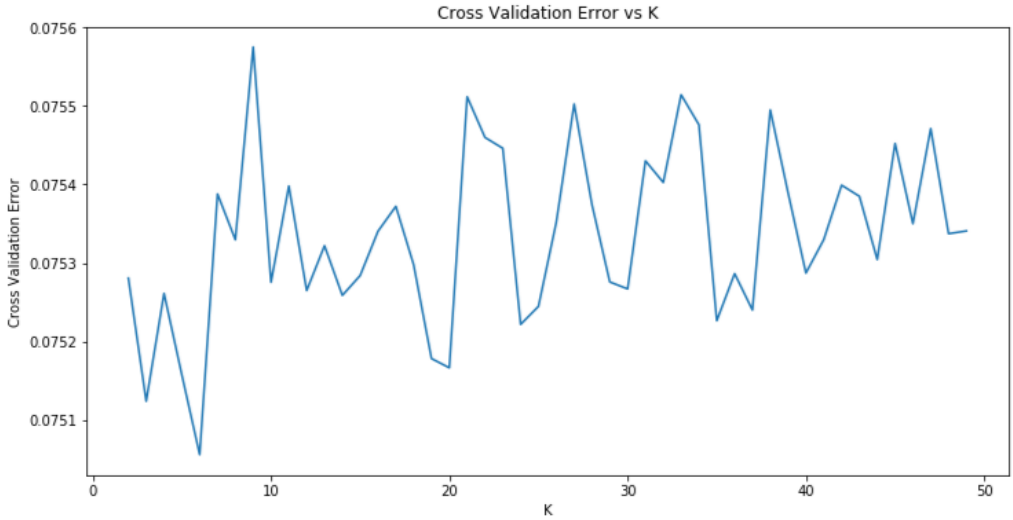
- In case of Lasso Regression, weights drastically drop to 0



We observe that ridge regression decreases the coefficients in a linear manner. In Lasso, weights converge to 0 more faster.

QUESTION 5
Implement k-fold cross validation. Analyse how behavior changes with different values of k.

As the value of k increases, number of training samples in each fold decreases. This leads to overfitting of the model. Also, computation time is directly proportional to value of k.



K varies from 2 to 50. $\alpha = 0.001$, $\lambda = 0.1$