**Statistical Learning Lab: Assignment-04**

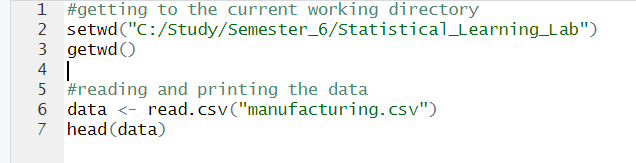
**Cross Validation and Bootstrapping**

**Name: Semanti Ghosh**

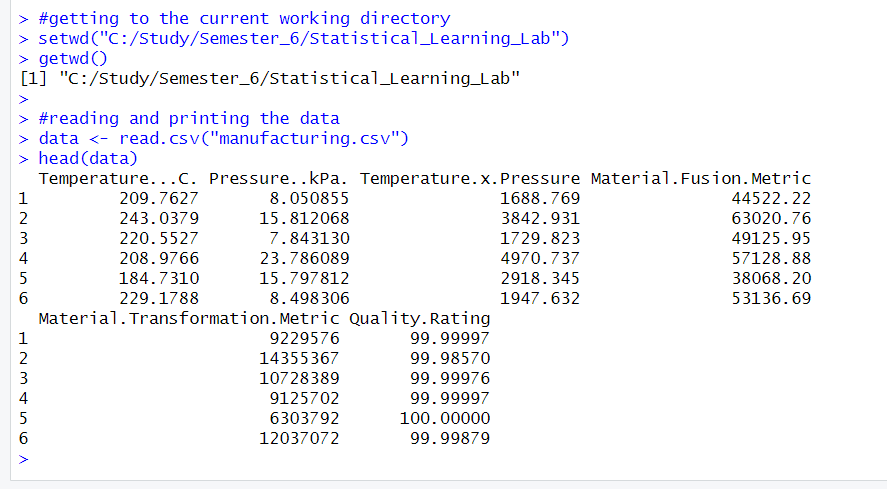
**Roll No: 22IM10036**

1. **Loading the data and printing the first few lines**

Code snippet

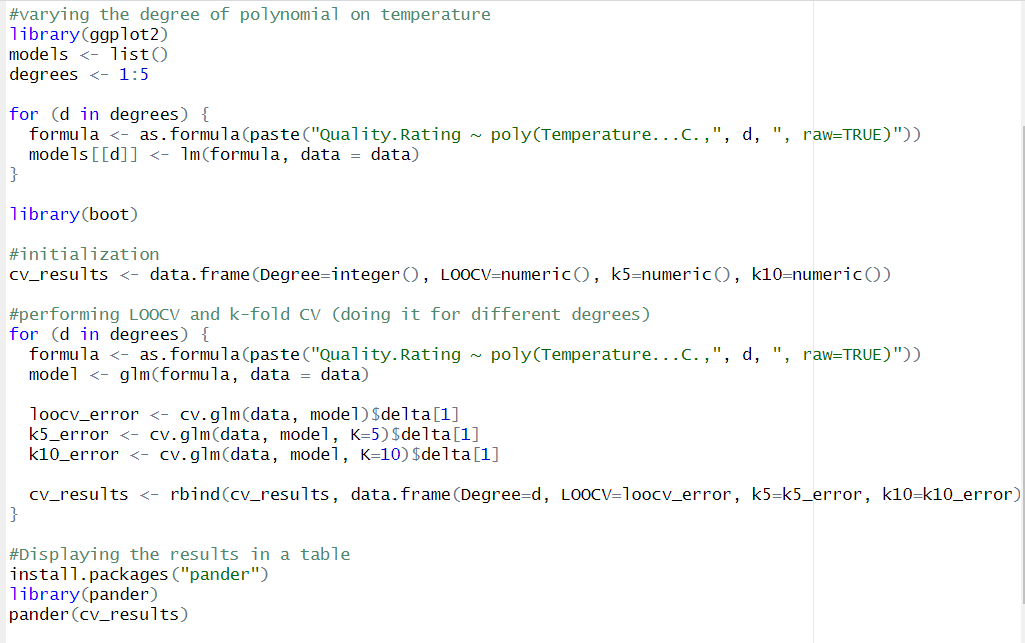


Output

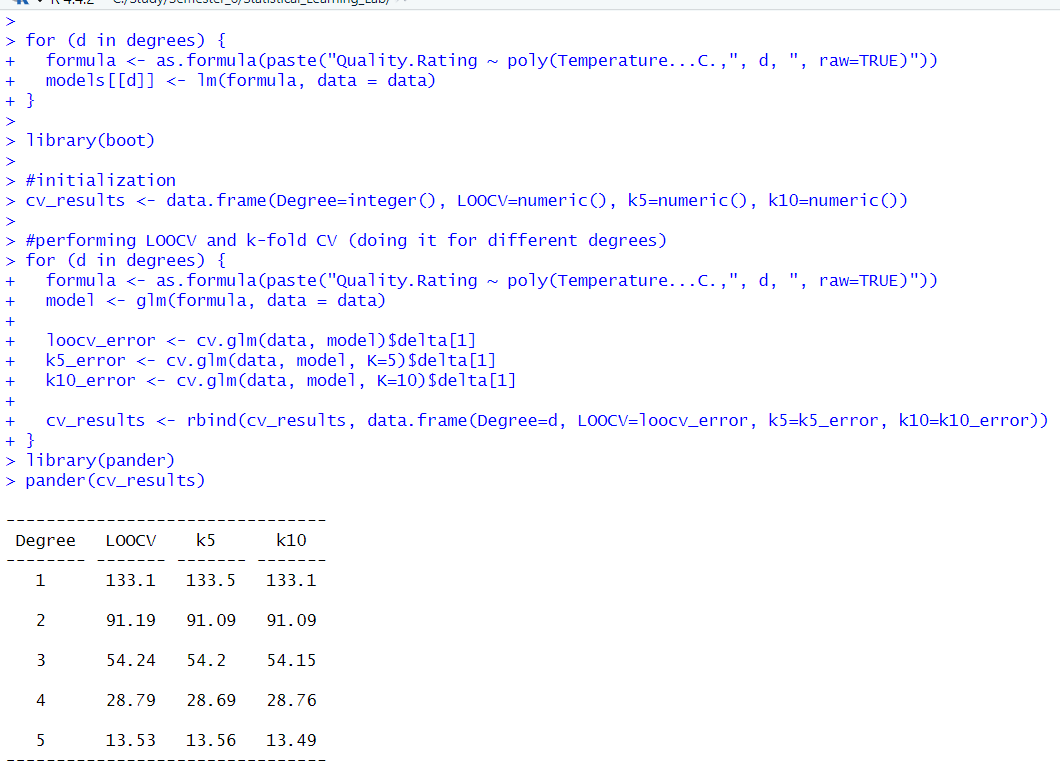


1. **Fitting polynomials on temperature from 1 to 5 degrees and then performing LOOCV and k-Fold CV**

Code Snippet

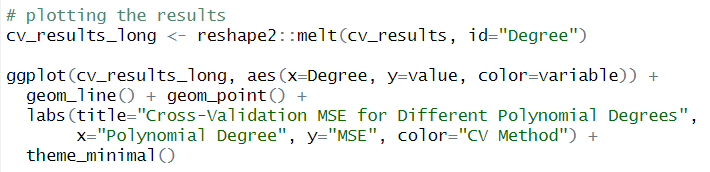


Output

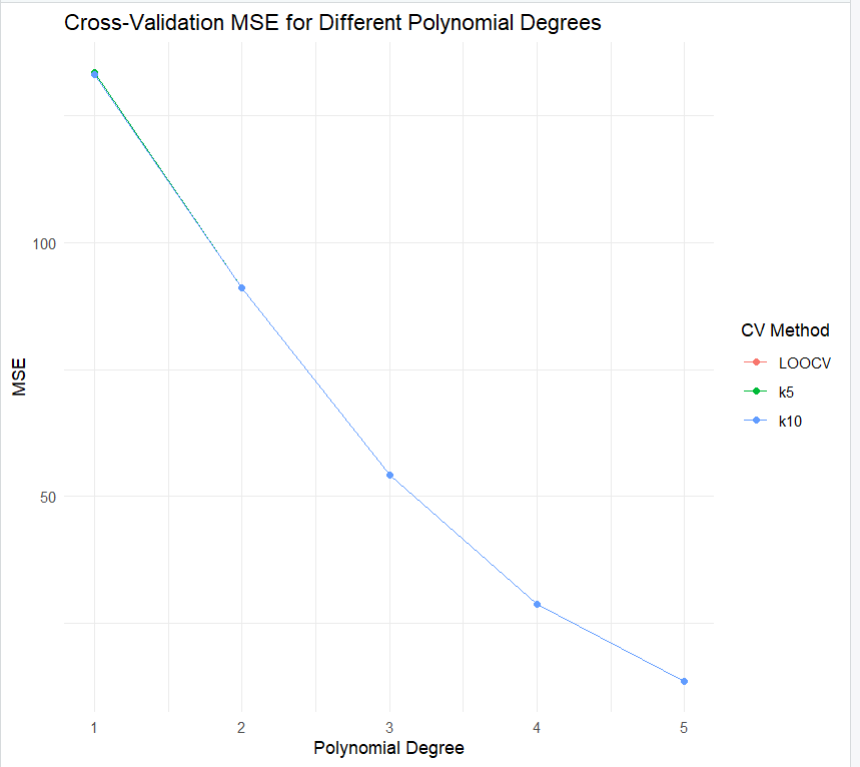


Plotting the results

Code snippet



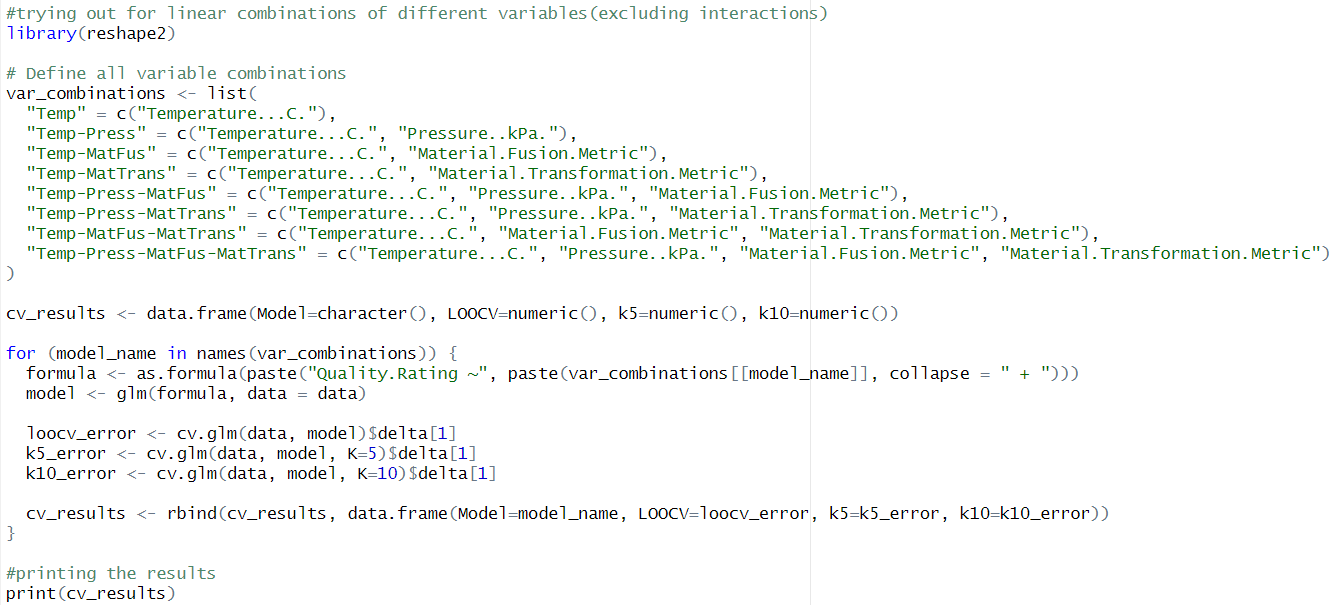
Output



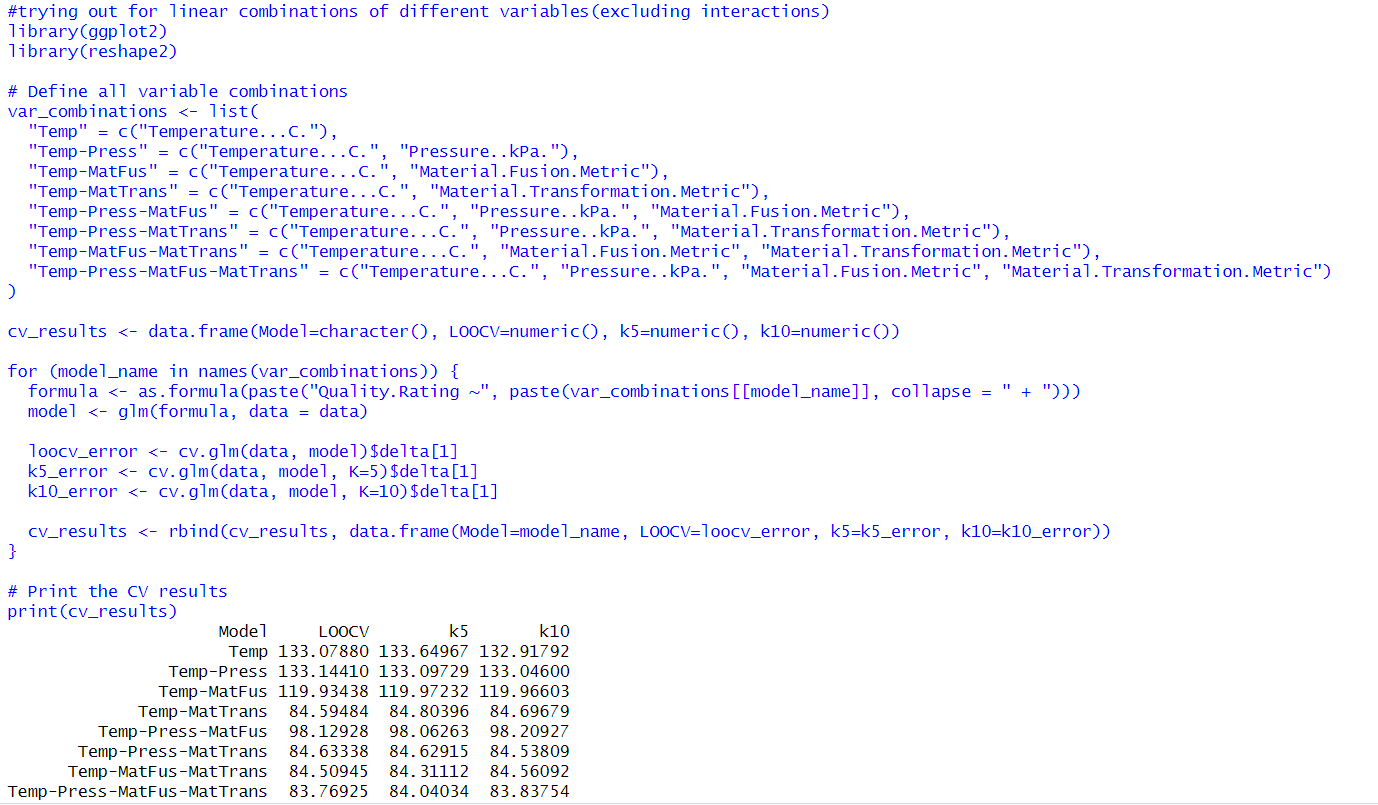
**Analysis:** From the above graph and results, we can say that when we fit a polynomial of degree 5, then the MSE is minimised. This is expected since the training and validation set error does decrease when the degree of the polynomial increases, since then the polynomial becomes more flexible. However, we should also keep in mind that if the degree of the polynomial is too high, then it will overfit the data, resulting in high variance. So, while the results suggest that 5 is the best choice, we might choose a lower degree polynomial (like maybe three or four degree polynomial). This will make the model more robust and ensure generalisation on unseen data.

1. **Fitting linear polynomials for different variables**

Code Snippet

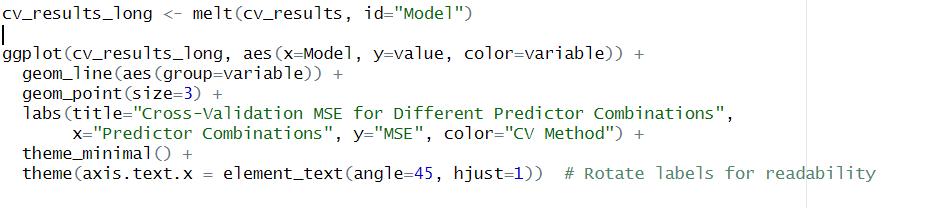


Output

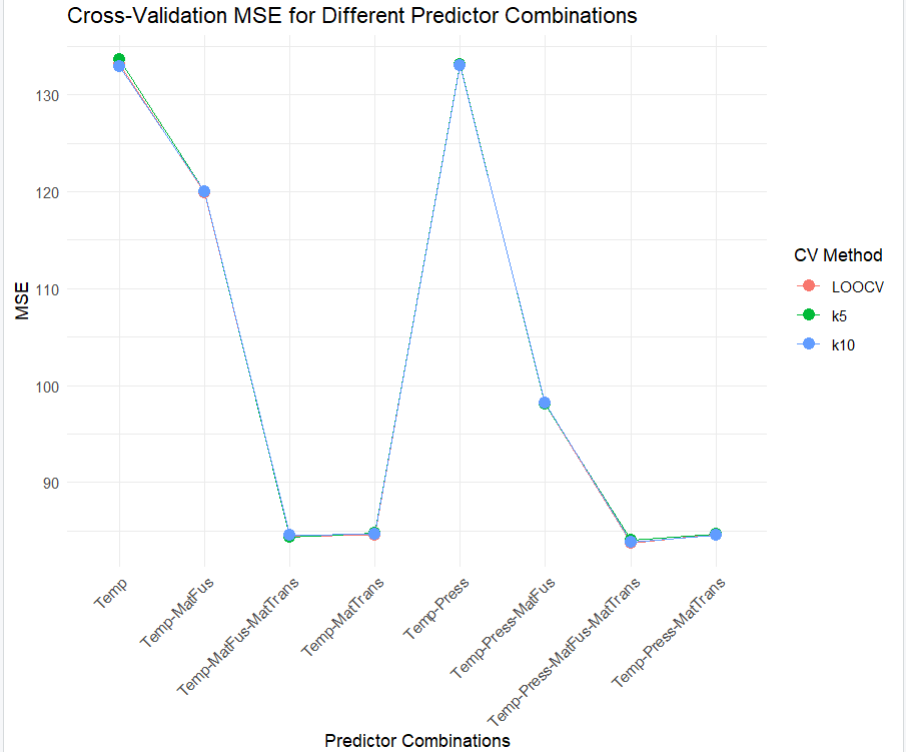


Plotting the results

Code Snippet



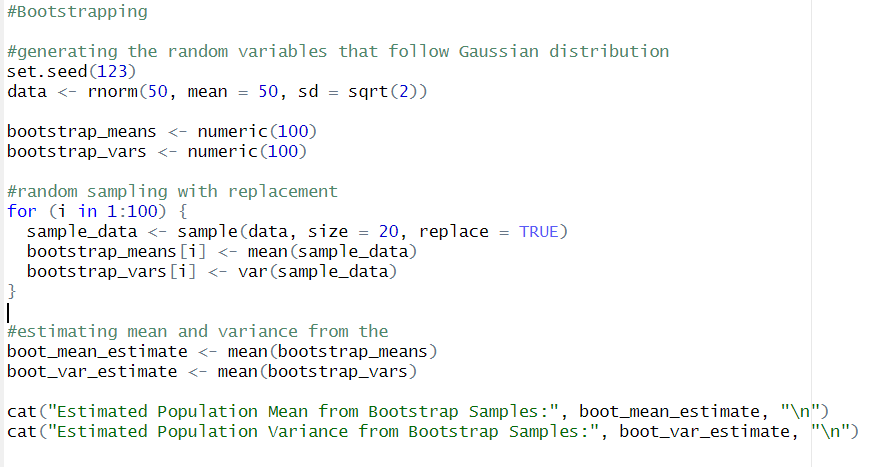
Output



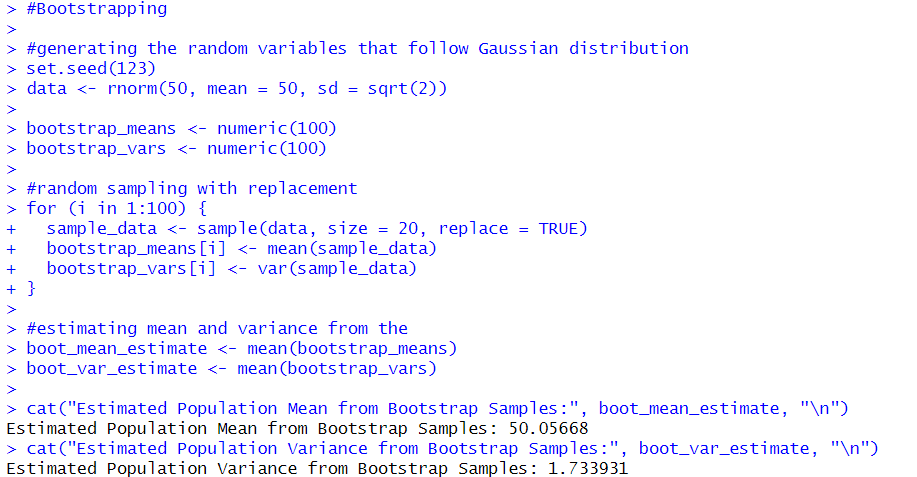
**Analysis**: From the results obtained, we can say that the error is minimised when all four parameters (Temperature, Pressure, Material Fusion Metric and Material Transformation Metric) are taken into consideration. However, like in the previous case, this can lead to overfitting resulting in high variance. So, even if the error is slightly greater, it’s safer to go for three parameters, like in Temperature, Material Fusion Metric, Material Transformation Metric. This will make the model more robust and ensure generalisation on unseen data.

1. **Bootstrapping**

Code snippet



Output



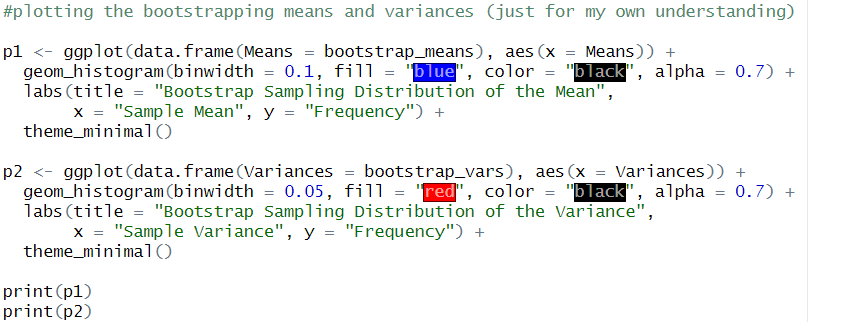
**Result**

**Estimated Population Mean = 50.05668**

**Estimated Population Variance = 1.733931**

Plotting the graph of means and variances in bootstrapping for better visualisation

Code Snippet



Output

