Due 11:59 PM CT 06/11/2023

Consider a well-known dataset on per capita income and per capita spending in public schools by state in the United States in 1979. (Available on blackboard). This dataset has been widely analyzed in various statistical. As in those previous analyses, we take per capita spending (Expenditure) as the dependent variable and per capita income as the predictor variable.

1. Draw a scatter-plot to check the relationship between Income and Expenditure and interpret the relationship between Income and Expenditure.
2. Find and interpret the slope for the least squares regression line
3. Find and interpret y-intercept for the least squares regression line
4. Find the least square regression equation and circle the results from your outputs.
5. Find proportion of the variation that can be explained by the least squares regression line (i.e., R2).
6. Find the estimator of σ2 (i.e., s2) and interpret the value of this estimator.
7. Check if the data contain any outlier or influential points?
8. Fit a single linear model and conduct 10-fold CV to estimate the error. In addition, draw the scatter plot with the fitted line and the scatter plot between the observed and fitted values below.

1. Fit a quadratic model and conduct 10-fold CV to estimate the error and draw the scatter plot with the fitted line and the scatter plot between the observed and fitted values.
2. Fit a mars model with optimal tuning parameters that you choose and conduct 10-fold CV to estimate the error and draw the scatter plot with the fitted line and the scatter plot between the observed and fitted values.
3. Compare the three fitted models in terms of RMSE and R2, and then make a recommendation based on your criteria.