

CS 5565 Virtual presentation

Annual Rainfall Prediction

Introduction



- To predict the annual rainfall of Missouri state based on past years (1883 - 2016) weather data of Missouri and its neighboring states which share most of the boundaries such as Iowa, Kansas, Illinois and Arkansas. The dataset is obtained from the NOAA website and data cleanup is done to combine all monthly observations in one dataset.
- The dataset that we have taken has 10 predictors [Year, month, latitude, longitude, elevation, TMAX, TMIN, TOBS, SNOW, SNWD] which can be used to predict the rainfall. We have used best subset selection method along with cross validation to select the most relevant subset of predictors. Best subset selection method has given lowest cv error for 10 predictors selection, however as we have observed the cv errors for all the subset predictor models, the cv errors are consistent for predictors beyond 6 (Figure 1) hence we have considered the model with best 6 predictors subset i.e. Month, station Latitude, station Elevation, max temp, min temp, snow.
- By selecting 6 predictors we are avoiding overfitting of predictors in model where they do not reduce CV errors significantly. This also achieves simpler model because high influence predictors are used.

Best Subset Selection

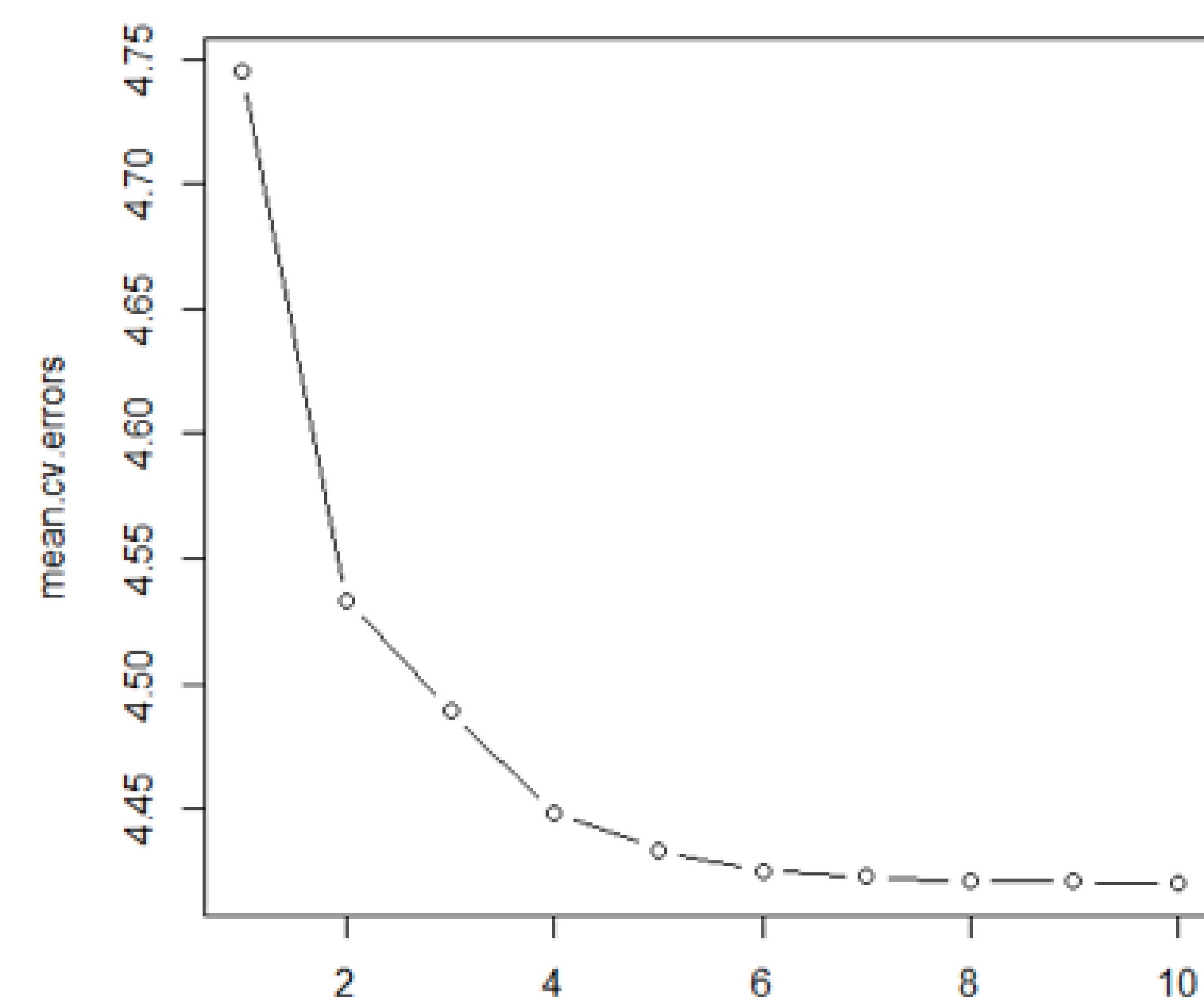


Figure 1

Method

- KNN Regression is a regression model which predicts the response variable based on the K nearest neighbors observations.
- We choose KNN regression because it is good fit for our use case where the annual rainfall prediction of Missouri state depends climate conditions of Missouri state and its neighboring states. We have selected neighboring states because we understood that there is a correlation between climate conditions of adjacent regions and it follows the KNN methodology to predict region's precipitation level by averaging precipitation levels of surrounding region's data.
- Our dataset consists records from 1883 to 2016. We are training our model by considering data of 1883-2014 as training data and we are predicting the annual rainfall for 2015 and 2016.
- We understand that precipitation has strong correlation with temperature and many other climate conditions and predicting rainfall would be reliable if temperature and other climate conditions are predicted with acceptable accuracy.

True Response vs Predicted Response

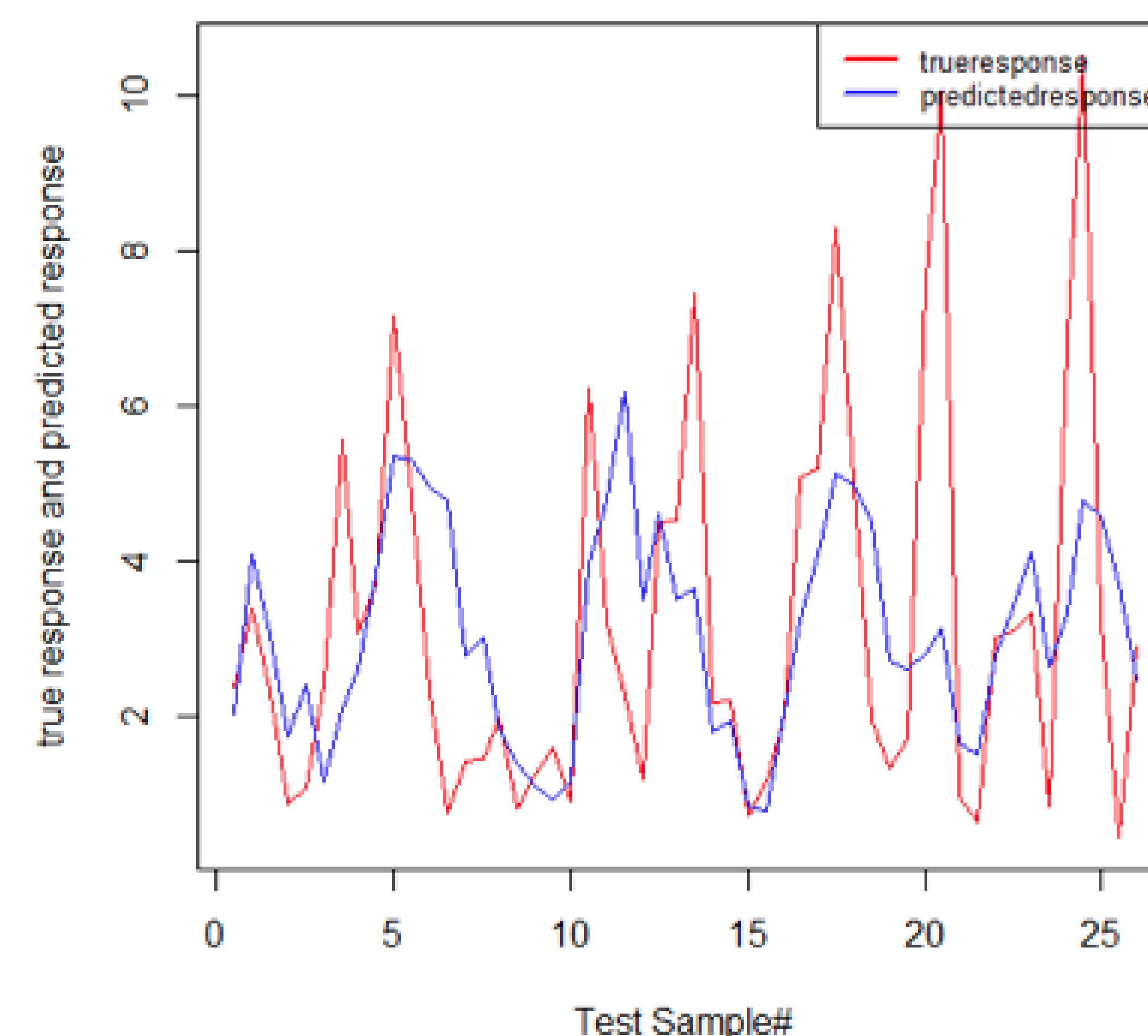


Figure 2

- Choosing value of K :** It is important to choose a optimal value of K to predict results with high accuracy and to find such optimal value of K, we used knn.reg() function to yield R^2 for multiple values of K without using the test data. These results were compared and we found value of K as 38 being optimal based on R^2 value for prediction.

Results Evaluation

- The plot in figure 2 shows are true response and predicted response for 52 test observations.
- From the graph we can see that predicted response follows the true response movement along the x-axis. This shows that predicted response is quantifying all the predictors in correct way as expected by the KNN regression model.
- The difference between predicted value and true response value could be caused because of -ve values of TMIN and thereby creating a negative coefficient in the model, however we feel that minimum temperatures don't have much impact on precipitation.

Conclusion & Future work

- The data cleanup is the major activity before starting analysis of dataset and same is the case with this project.
- Once the data is in correct format, we are able to generate KNN regression model and fit huge training dataset that ranged over 144 years.
- As shown in the results, we can conclude that KNN is a best fit for the problems of these kind where prediction is in accordance with the observations around certain region.
- We plan on including more neighboring states and see the results because the wind directions over certain states could be impactful in precipitation of certain region hence covering all the directions around region would be a possible next upgrade for prediction accuracy improvement.

References

- <http://artax.karlin.mff.cuni.cz/r-help/library/FNN/html/knn.reg.html>
- <https://www.ncdc.noaa.gov/oa/climate/ghcn-daily/>
- <ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/>
- <http://www.nwmissouri.edu/international/northwest.htm>