46886: Machine Learning Fundamentals 23-24 A3

Homework 1

1/23/2024

Team 6

Megha Maliwal , Sherine George George , Shashank Kallahalli Suresh

(a) Build a linear regression model to predict the dependent variable Temp, using CO2, CH4, N2O, CFC-11, CFC-12, Aerosols, TSI and MEI as features (Year and Month should NOT be used as features in the model). As always, use only the training set to train your model. What are the in-sample and out-of-sample R2 , MSE, and MAE?
Answer:
In-sample R2 is: 0.6920595959984911
In-sample MSE is: 0.008731426409910696
In-sample MAE is: 0.07260918612938558
Out-of-sample R2 is: -0.5413255834033228
Out-of-sample MSE is: 0.012206974835145303
Out-of-sample MAE is: 0.09312747891279667

(b) Build another linear regression model, this time with only N2O, Aerosols, TSI, and MEI as features. What are the in-sample and out-of-sample R2, MSE, and MAE?
Answer:
In-sample R2 is: 0.6490120806760418
In-sample MSE is: 0.009952007429105654
In-sample MAE is: 0.07666650280233073
Out-of-sample R2 is: 0.2003186110455797
Out-of-sample MSE is: 0.006333308611893898
Out-of-sample MAE is: 0.06154027269393324

(c) Between the two models built in parts (a) and (b), which performs better in-sample? Which performs better out-of-sample?
Answer: Between the two models, the model built in part (a) performs better in-sample, while the model built in part (b) performs better out-of-sample. This can be easily seen keeping in mind a higher R2 is better, and a lower MSE and MAE is better.

(d) For each of the two models built in parts (a) and (b), what was the regression coefficient for the N2O feature, and how should this coefficient be interpreted?
Answer: The regression coefficient for the N2O feature in part(a) was:
-0.034847807455777306
A unit change in NO2 (i.e., an increase of 1 ppmv in NO2) leads to a change of -0.0348 in the difference (in degrees Celsius) between the average global temperature in that period and a reference value.

The regression coefficient for the N2O feature in part(b) was: 0.024276120902004254
A unit change in NO2 (i.e., an increase of 1 ppmv in NO2) leads to a change of 0.0243 in the difference (in degrees Celsius) between the average global temperature in that period and a reference value.

(e) Given your responses to parts (c) and (d), which of the two models should you prefer to use moving forward?
Answer: Given our responses to parts (c) and (d), we should prefer the second model, i.e. the model built in part(b) moving forward. The current scientific opinion is that N2O is a greenhouse gas – a higher concentration traps more heat from the sun, and thus contributes to the heating of the Earth. This implies that an increase in NO2 concentration should increase the average global temperature, which means an increase in the 'Temp' variable. This means we should expect a positive regression coefficient for the N2O feature in the Linear Regression Model. We see this in the second model.