Estimated time to complete 1-2 hours

STUDYING Climate Change


There have been many studies documenting that the average global temperature has been increasing over the last century. The consequences of a continued rise in global temperature will be dire. Rising sea levels and an increased frequency of extreme weather events will affect billions of people.

In this problem, we will attempt to study the relationship between average global temperature and several other factors.

The file **climate_change.csv** contains climate data from May 1983 to December 2008. The available variables include:

- *Year*: the observation year.
- *Month*: the observation month.
- *Temp*: the difference in degrees Celsius between the average global temperature in that period and a reference value. **This is the Dependent variable.**
- *CO2*, *N2O*, *CH4*, *CFC.11*, *CFC.12*: atmospheric concentrations of carbon dioxide (CO_2), nitrous oxide (N_2O), methane (CH_4), trichlorofluoromethane (CCl_3F; commonly referred to as CFC-11) and dichlorodifluoromethane (CCl_2F_2; commonly referred to as CFC-12), respectively.
- CO2, N2O and CH4 are expressed in ppmv (parts per million by volume -- i.e., 397 ppmv of CO2 means that CO2 constitutes 397 millionths of the total volume of the atmosphere)
- CFC.11 and CFC.12 are expressed in ppbv (parts per billion by volume).

- *Aerosols*: the mean stratospheric aerosol optical depth at 550 nm. This variable is linked to volcanoes, as volcanic eruptions result in new particles being added to the atmosphere, which affect how much of the sun's energy is reflected back into space.
- *TSI*: the total solar irradiance (TSI) in W/m^2 (the rate at which the sun's energy is deposited per unit area). Due to sunspots and other solar phenomena, the amount of energy that is given off by the sun varies substantially with time.
- *MEI*: multivariate El Nino Southern Oscillation index (MEI), a measure of the strength of the El Nino/La Nina-Southern Oscillation (a weather effect in the Pacific Ocean that affects global temperatures).

Problem 1.1 - Creating Our First Model

We are interested in how changes in these variables affect future temperatures, as well as how well these variables explain temperature changes so far. To do this, first read the dataset climate_change.csv into Python.

1) Compute the correlations between all the variables in the dataset except Date & Month. Which of the following independent variables is N2O highly correlated with (absolute correlation greater than 0.7)?
2) Find out which variables are highly correlated with the Dependent Variable?
3) What will be the prediction of a baseline Model? What will be its SSE, more commonly known as SST/TSS?

4) Next, build a linear regression model to predict the dependent variable Temp, using CO2, N2O, CFC.12, and Aerosols as independent variables. Find the RMSE, SSE and MAPE for the above Model. What is the model R2 (the "Multiple R-squared" value & Adjusted R2).Calculate R2 value manually using the formulae discussed in class. Both should match.
5) **Find the correlation between the actual & predicted values of Temp. Square this number and you should get the Multiple R2. This is another quick way to compute R2. (R2= correlation(actual, predicted values)^2)**

Feel free to discuss the questions in the Group.