Assignment #3

Notice

Due Date: June 25, 24:00

You can submit your code in a format of .py file or jupyter-notebook file.

Total Points = 100 pts

---

1. File and Data Manipulation: Global Warming (50 pts)

In this problem, we will practice
(1) how to read a text file,
(2) how to manipulate data in the file,
(3) how to draw some figure to illustrate the data.

If you visit [https://climate.nasa.gov/vital-signs/ (https://climate.nasa.gov/vital-signs/)](https://climate.nasa.gov/vital-signs/), you can see statistical data indicating Global Warming, such as Carbon Dioxide Level, Global Temperature, and so on. You can download suc statistical data from the NASA website. In this exercise, we will read (1)Global Temperature and (2)Carbon Dioxide data from files, manipulate these data, and finally draw some graphs to visualize them.

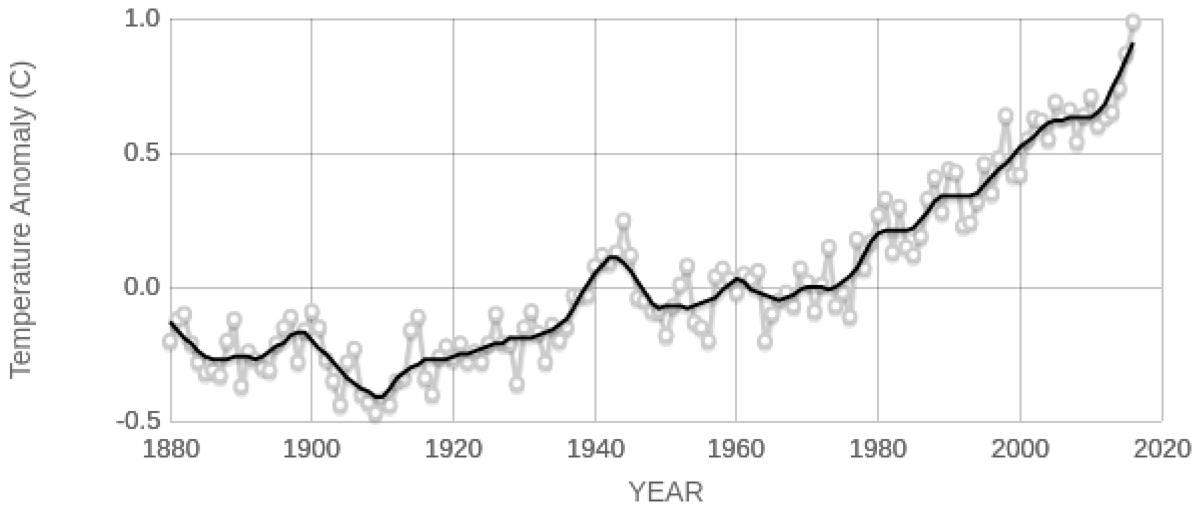
(1) Global Temperature (25 pts)

You can download Global Temperature data from [https://climate.nasa.gov/vital-signs/global-temperature/ (https://climate.nasa.gov/vital-signs/global-temperature/)](https://climate.nasa.gov/vital-signs/global-temperature/) whose file name is "647_Global_Temperature_Data_File.txt." If you open it, it reads

```
1880 -0.20 -0.13
1881 -0.12 -0.16
1882 -0.10 -0.19
1883 -0.21 -0.21
1884 -0.28 -0.24
1885 -0.32 -0.26
1886 -0.31 -0.27
...
```

The first column is the year where the global temperature was measured, which spans from 1880 to 2016.
The second column indicates the annual mean surface temperature relative to 1951-1980 average temperature.
The third column is the five-year mean of the surface temperature.

These data can be plotted as follows:



Source: climate.nasa.gov

Your homework! (25 points)

Please read the global temperature data of "647_Global_Temperature_Data_File.txt" and plot them that looks like the above figure.

---

(2) Carbon Dioxide (25 pts)

You can download Carbon Dioxide data from https://climate.nasa.gov/vital-signs/carbon-dioxide/ (https://climate.nasa.gov/vital-signs/carbon-dioxide/) whose file name is "co2_mm_mlo.txt." This file consists of two parts. The first part is the description how to use the data contained in the file, which starts with # at each line. The second part is Carbon Dioxide data, which we want to use. The second part reads

```
1958   3    1958.208      315.71      315.71      314.62     -1
1958   4    1958.292      317.45      317.45      315.29     -1
1958   5    1958.375      317.50      317.50      314.71     -1
1958   6    1958.458      -99.99      317.10      314.85     -1
1958   7    1958.542      315.86      315.86      314.98     -1
1958   8    1958.625      314.93      314.93      315.94     -1
1958   9    1958.708      313.20      313.20      315.91     -1
1958  10    1958.792      -99.99      312.66      315.61     -1
1958  11    1958.875      313.33      313.33      315.31     -1
1958  12    1958.958      314.67      314.67      315.61     -1
1959   1    1959.042      315.62      315.62      315.70     -1
1959   2    1959.125      316.38      316.38      315.88     -1
1959   3    1959.208      316.71      316.71      315.62     -1
1959   4    1959.292      317.72      317.72      315.56     -1
1959   5    1959.375      318.29      318.29      315.50     -1
```

The first and second columns are the year and the month where the carbon dioxide level was measured.
The fourth column contains "the monthly mean CO2 mole fraction determined from daily averages." as stated in the description. We do not need remaining columns for our purpose.

There is one thing that we should be careful about: "-99.99" in the fourth column means "Missing data.", but is not a measured carbon dioxide level. When we do manipulate data in the fourth column, we should exclude "-99.99" in your data.

Your homework! (25 points)

1. Please read carbon dioxide level data of "co2_mm_mlo.txt"

2. Plot carbon dioxide levels at one-month intervals. One month is $1/12 = 0.83333 \cdots \text{year}$. For example, October 1958 is 1958.83333 [year]. If the carbon dioxide level is "-99.99", then skip the corresponding month.

3. Calculate the yearly mean carbon dioxide mole fractions from the monthly mean data.
   For example, the yearly mean carbon dioxide mole fraction in 1958 is
   $$(315.71 + 317.45 + 317.50 + 315.86 + 314.93 + 313.20 + 313.33 + 314.67)/8 = 315.33125$$

4. Plot the yearly mean carbon dioxide mole fraction data together with the monthly mean data read from "co2_mm_mlo.txt."

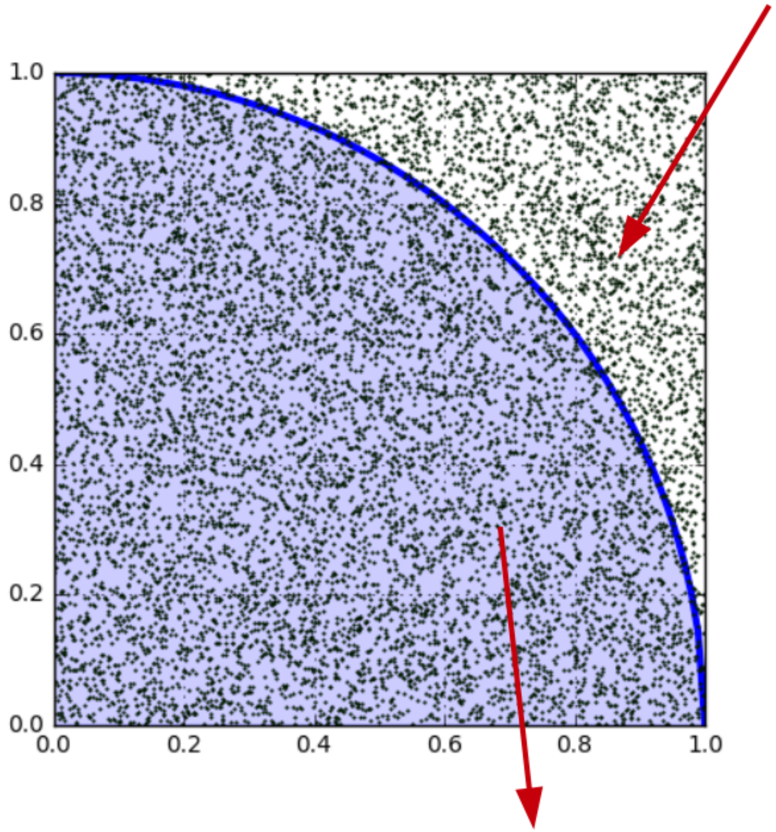
---

2. Estimate π Number by using the Monte-Carlo Integration (50 pts)

We have learned the Monte-Carlo Integration by drawing points randomly generated. Here we will apply the Monte-Carlo integration scheme to estimate the π number.

---

1. Please randomly generate n_{tot} pairs of (x, y) inside the square region, $0 \leq x \leq 1, \ \ 0 \leq y \leq 1$



2. Among n_{tot} pairs, count the number of pairs located inside the quater of the unit circle, $x^2 + y^2 \leq 1$
   The number of counted pairs is denoted by n.

3. π can be estimated as $4\dfrac{n}{n_{tot}}$: $\ \ \pi \approx 4\dfrac{n}{n_{tot}}$

The Monte-Carlo estimation can be done as follows:

1. Prepare the region $S = \{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq 1\}$, which is the square whose side length is 1.

2. Generate pairs (x_i, y_i) located inside $S = \{(x, y)|0 \leq x \leq 1, 0 \leq y \leq 1\}$ as many as you want by using the random number generator. The total number of points randomly generated is n_{tot}.

3. Among randomly generated pairs, count the number of points inside a circle $x^2 + y^2 = 1$, which is denoted by n.

4. The area of a quater of the circle in S is given by $\pi/4$. According to the Monte-Carlo integration, The area of a quater of the circle in S is estimated by $\frac{n}{n_{tot}}$. Thus π is approximated as

$$\pi \cong 4\frac{n}{n_{tot}}$$

Your homework! (50 pts)

1. Following the steps described above, estimate the π number and plot the above graph with your random sampling points.
2. **(Extra: 20 pts)** Can you repeat the above estimation many times, for example, 100 times? Calculate the average of estimations and the corresponding standard deviation.

---

3. (Optional) Curve Fitting to Random Data Points (50 pts)

This problem is not required, but optional. If you want to get more points, please solve these problems.

We have learned the curve-fitting algorithm in order to find out a function that can explain the trend of random data points.

1. Linear Regression (Linear Curve-Fitting)
2. Linearization of Variable Transformation
3. Polynomial Regression (Non-linear Curve-Fitting)

Please apply the above methods to find out a function best fit to given random data points.

Your homework! (50 points)

Find out a function best fit to data points in "curvefit-data1.dat".