Summer 2024

MATH 4650-E01
CSCI 4650-E01
MATH 5660-E01 (additional exercises will be assigned by email for 5660 students)

Homework 5

Due on Monday July 8th 2024

Exercises:

1. EX.4.1.1.a handwritten or typed
2. EX.4.1.1.c handwritten or typed
3. EX.4.1.2.a handwritten or typed
4. EX.4.1.8.b handwritten or typed
5. EX.4.1.9.b handwritten or typed
6. CP.4.1.5 colab
7. CP.4.2.6 handwritten or typed
8. CP.4.2.7 handwritten or typed
9. CP.4.2.8 handwritten or typed
10. CP.4.2.9 colab+(handwritten or typed)
11. EX.4.3.2.b handwritten or typed
12. EX.4.3.7.b handwritten or typed
13. EX.4.3.8.b handwritten or typed
14. CP.4.3.6 colab
15. CP.4.5.1.b colab
16. CP.4.5.2.b colab
17. CP.4.5.3.b colab
18. CP.4.5.4.b colab
19. CP.4.5.8 colab
20. CP.4.5.11.b colab

EX.4.1.1.a, Sauer3

Solve the normal equations to find the least squares solution and 2-norm error for the following inconsistent system

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}.$$

EX.4.1.1.c, Sauer3

Solve the normal equations to find the least squares solution and 2-norm error for the following inconsistent system

$$\begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 2 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 2 \end{bmatrix}.$$

EX.4.1.2.a, Sauer3

Find the least squares solutions and RMSE of the following systems:

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

EX.4.1.8.b, Sauer3

Find the best line through each set of data points, and find the RMSE

$$(1, 2), \quad (3, 2), \quad (4, 1), \quad (6, 3).$$

EX.4.1.9.b, Sauer3

Find the best parabola through each data point set in EX.4.1.8.b, and compare the RMSE with the best-line fit. The points are

$$(1, 2), \quad (3, 2), \quad (4, 1), \quad (6, 3).$$

Instructions:

Please handwrite the matrix A and b, and then use python to solve the linear least squares problem by giving x to 3 digits after the dot. Then write the best-fit polynomial (with 3 digits after the dot coefficients). Then compute the RMSE with Python and please compare the RMSE of this problem with the RMSE of EX.4.1.8.a.

So to repeat, I need: (1) A, b, (2) x, (3) parabola, (4) RMSE, and (5) comparison of RMSEs.

CP.4.1.5, Sauer3

A company test-markets a new soft drink in 22 cities of approximately equal size. The selling price (in dollars) and the number sold per week in the cities are listed as follows:

city	price	sales/week
1	0.59	3980
2	0.80	2200
3	0.95	1850
4	0.45	6100
5	0.79	2100
6	0.99	1700
7	0.90	2000
8	0.65	4200
9	0.79	2440
10	0.69	3300
11	0.79	2300
12	0.49	6000
13	1.09	1190
14	0.95	1960
15	0.79	2760
16	0.65	4330
17	0.45	6960
18	0.60	4160
19	0.89	1990
20	0.79	2860
21	0.99	1920
22	0.85	2160

a. First, the company wants to find the “demand curve”: how many it will sell at each potential price. Let P denote price and S denote sales per week. Find the line $S = c_1 + c_2 P$ that best fits the data from the table in the sense of least squares. Find the normal equations and the coefficients c_1 and c_2 of the least squares line. Plot the least squares line along with the data, and calculate the root mean square error.

b. After studying the results of the test marketing, the company will set a single selling price P throughout the country. Given a manufacturing cost of \$0.23 per unit, the total profit (per city, per week) is $S(P - 0.23)$ dollars. Use the results of the preceding least squares approximation to find the selling price for which the company's profit will be maximized.

Getting started: https://colab.research.google.com/drive/13gw6Vwiz6cQjx3lElAJtqnxC6vhabWkv

CP.4.2.6, Sauer3

The bloodstream concentration of a drug, measured hourly after administration, is given in the accompanying table. Fit the model

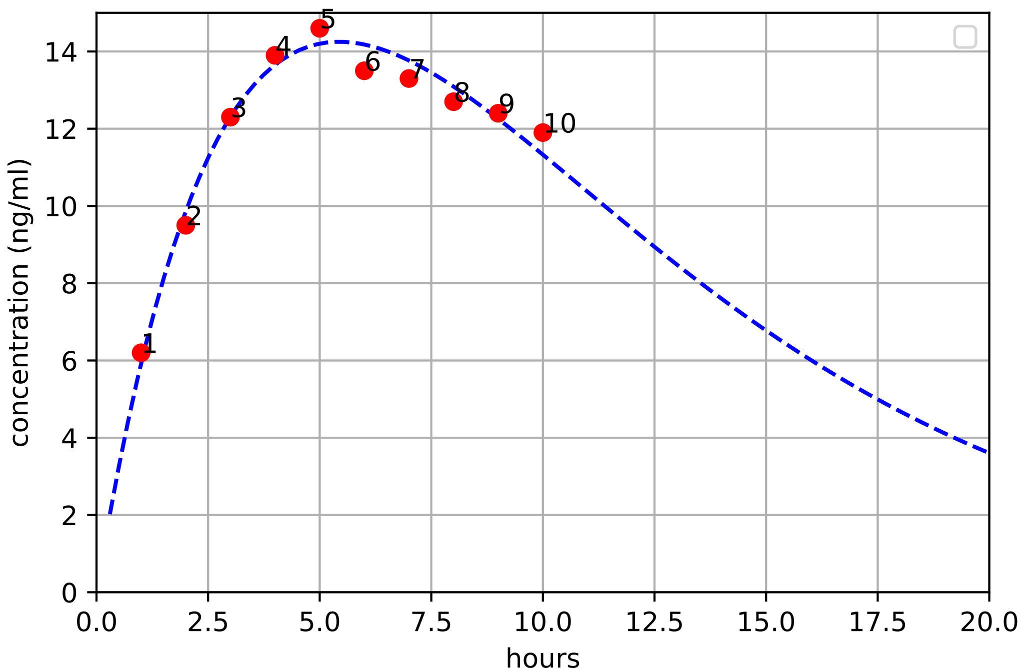
$$y = c_1 t e^{c_2 t} \qquad (4.21).$$

Find the estimated maximum and the half-life. Suppose that the therapeutic range for the drug is 4–15 ng/ml. Use the equation solver of your choice to estimate the time the drug concentration stays within therapeutic levels.

Do not worry about the questions of Sauer. This problem is (more or less) solved in the following notebook. You have 9 questions to answer next page.

hours	concentration (ng/ml)
1	6.2
2	9.5
3	12.3
4	13.9
5	14.6
6	13.5
7	13.3
8	12.7
9	12.4
10	11.9

Jupyter Notebook: https://colab.research.google.com/drive/1KQChAXtFm-tQmeb9GvaHJPkkNDIiarVY



Handwritten questions:

a. In the code what do the array `xx` and the array `yy` represent?

b. Explain what this line do:
```
A = np.array([ np.ones( xx.shape ), xx ]).T
```
Note: Writing "this creates the matrix A" is not enough. Something like "this creates the matrix A such as the first column is ... and the second column is ..., each row represents ..." is needed

c. Explain what this line do:
```
b = np.array([ np.log(yy) - np.log(xx) ]).T
```

d. Explain what this line do: `x = np.linalg.lstsq(A,b,rcond=None)[0]`

e. Explain what this line do: `yyy[i] = exp(x[0]) * xxx[i] * exp( x[1] * xxx[i] )`

f. We find that the model is

$$c(t) = 7.12\ t\ e^{-0.18t}.$$

What is $c(t)$? What is t?

g. Do you think we have a good fit? Is this a good model?

h. Describe biologically/physically what is happening? Why is the curve rapidly increasing, and then slowly decreasing with time?

i. Visually, answer the question: "Suppose that the therapeutic range for the drug is 4–15 ng/ml. Estimate the time interval the drug concentration stays within therapeutic levels." ("Visually" means "look at the graph and guess an interval".)

CP.4.2.7, Sauer3

The file `windmill.txt`, available from the textbook website, is a list of 60 numbers which represent the monthly megawatt-hours generated from Jan. 2005 to Dec. 2009 by a wind turbine owned by Minnkota Power Cooperative near Valley City, ND. The data is currently available at `http:www.minnkota.com`. For reference, a typical home uses around 1MWh per month.

a. Find a rough model of power output as a yearly periodic function. Fit the data to equation (4.9),

$$f(t) = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t + c_4 \cos 4\pi t$$

where the units of t are years, that is $0 \leq t \leq 5$, and write down the resulting function.

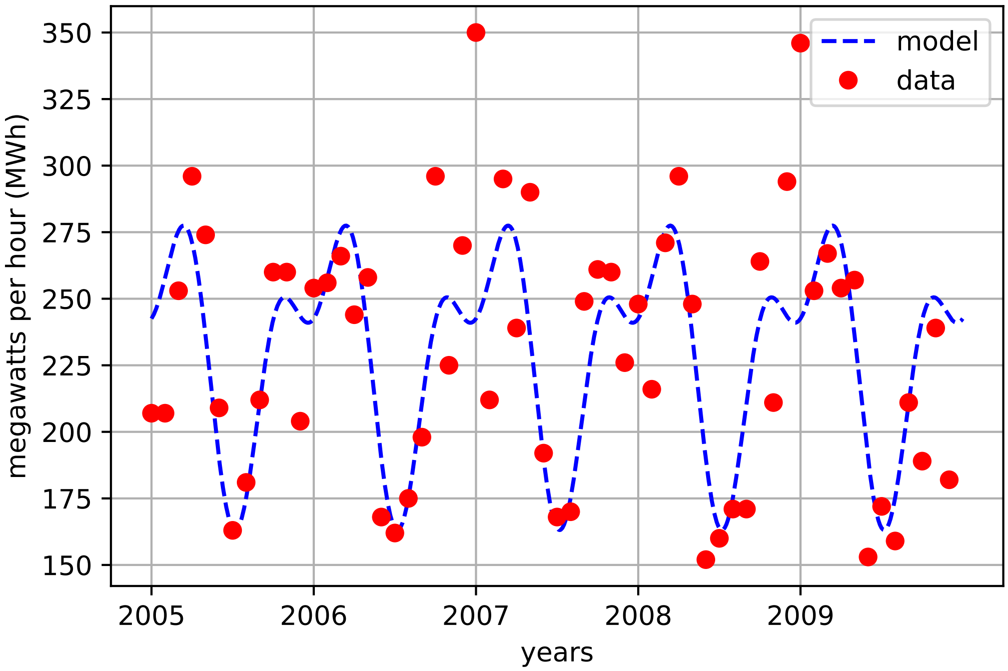
b. Plot the data and the model function for years $0 \leq t \leq 5$. What features of the data are captured by the model?

Do not worry about the questions of Sauer. This problem is (more or less) solved in the following notebook. You have 4 handwritten questions to answer next page.

	2005	2006	2007	2008	2009
Jan.	207	254	350	248	346
Feb.	207	256	212	216	253
Mar.	253	266	295	271	267
Apr.	296	244	239	296	254
May	274	258	290	248	257
Jun.	209	168	192	152	153
Jul.	163	162	168	160	172
Aug.	181	175	170	171	159
Sep.	212	198	249	171	211
Oct.	260	296	261	264	189
Nov.	260	225	260	211	239
Dec.	204	270	226	294	182

Production in megawatt-hours generated from Jan. 2005 to Dec. 2009.

Jupyter Notebook: https://colab.research.google.com/drive/1DJFK_hsx23JiX1GF4BFuaNKCtLG7Foe0



Handwritten questions:

a. In the code what does the array `t` represent?

b. Explain what these lines do:

```
A = np.zeros( [ 60, 4] )
A[:,0] = np.ones( 60)
A[:,1] = np.cos(2*pi*t)
A[:,2] = np.sin(2*pi*t)
A[:,3] = np.cos(4*pi*t)
```

Note: Writing "this creates the matrix A" is not enough. Something like "this creates the matrix A such as the first column is ... and the second column is ..., each row represents ..." is needed

c. We find that the model is

$$f(t) = 229.9 + 39.3 \cos 2\pi t + 14.5 \sin 2\pi t - 26.7 \cos 4\pi t$$

What is $f(t)$? What is t? Which month does $t = 1$ represent? What is the value of t that represents the month of April 2007?

d. Do you think we have a good fit? Is this a good model?

CP.4.2.8, Sauer3

The file `scrippsy.txt`, available from the textbook website, is a list of 50 numbers which represent the concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), recorded at Mauna Loa, Hawaii, each May 15 of the years 1961 to 2010. The data is part of a data collection effort initiated by Charles Keeling of the Scripps Oceanographic Institute (Keeling et al. [2001]). Subtract the background level of 279 ppv as in Computer Problem 4, and fit the data to an exponential model. Plot the data along with the best fit exponential function and report RMSE.

year	ppm	year	ppm	year	ppm	year	ppm	year	ppm
1961	320.58	1971	328.92	1981	342.91	1991	359.34	2001	374.02
1962	321.01	1972	330.07	1982	344.14	1992	359.66	2002	375.55
1963	322.25	1973	332.48	1983	345.75	1993	360.28	2003	378.35
1964	322.24	1974	333.09	1984	347.43	1994	361.68	2004	380.61
1965	322.16	1975	333.97	1985	348.93	1995	363.79	2005	382.24
1966	324.01	1976	334.87	1986	350.21	1996	365.41	2006	384.94
1967	325.00	1977	336.75	1987	351.84	1997	366.80	2007	386.43
1968	325.57	1978	338.01	1988	354.22	1998	369.30	2008	388.49
1969	327.34	1979	339.47	1989	355.67	1999	371.00	2009	390.18
1970	328.07	1980	341.46	1990	357.16	2000	371.82	2010	393.22

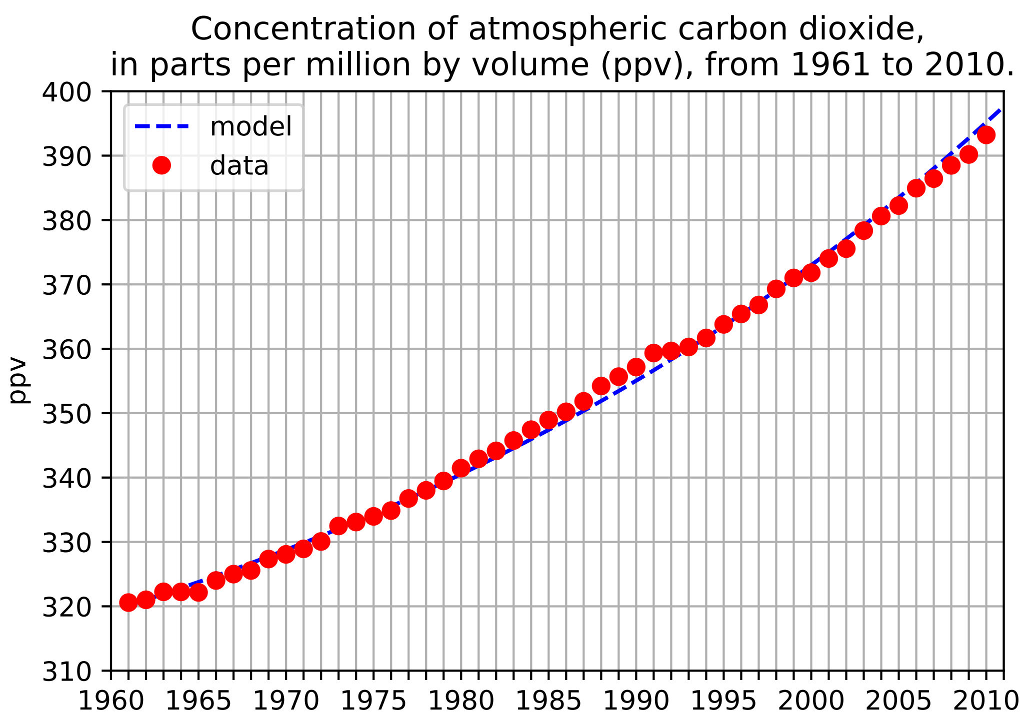
Concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), from 1961 to 2010.

See the website `http://scrippsco2.ucsd.edu` for much more data and analysis of the Scripps carbon dioxide study.

Do not worry about the questions of Sauer. This problem is (more or less) solved in the following notebook. You have 1 handwritten question to answer. See below.

Handwritten Question: By using `print`, give the equation of the best-fit exponential function. Be careful that there are two offsets. We offset the y-axis by 279 ppv, and we offset the year axis by 1961. Check that for year 1970, the best-fit value (given by your function) is 328.77 ppv.

Jupyter Notebook: https://colab.research.google.com/drive/15l78g98Fkzl65f_C2uqjC8WCBzK31qT-



CP.4.2.9, Sauer3

The file `scrippsm.txt`, available from the textbook website, is a list of 180 numbers which represent the concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), recorded monthly at Mauna Loa from Jan. 1996 to Dec. 2010, taken from the same Scripps study as Computer Problem 8.

a. Carry out a least squares fit of the CO_2 data using the model

$$f(t) = c_1 + c_2 t + c_3 \cos(2\pi t) + c_4 \sin(2\pi t)$$

where t is measured in months. Report the best fit coefficients c_i and the RMSE of the fit. Plot the continuous curve from Jan. 1989 to the end of this year, including the 180 data points in the plot.

b. Use your model to predict the CO_2 concentration in May 2004, Sept. 2004, May 2005, and Sept. 2005. These months tend to contain the yearly maxima and mimima of the CO_2 cycle. The actual recorded values are 380.63, 374.06, 382.45, and 376.73 ppv, respectively. Report the model error at these four points.

c. Add the extra term $c_5 \cos(4\pi t)$ and redo part (a) and (b). Compare the new RMSE and the four model errors.

d. Repeat part (c) using the extra term $c_5 t^2$. Which terms lead to more improvement in the model, part(c) or (d)?

e. Add both terms from (c) and (d) and redo parts (a) and (b). Prepare a table summarizing your results from all parts of the problem, and try to provide an explanation for the results.

See the website `http://scrippsco2.ucsd.edu` for much more data and analysis of the Scripps carbon dioxide study.

Do not worry about the questions of Sauer. This problem is (more or less) solved in the following notebook. You have one Colab question and one Handwritten question to answer. See below.

	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Jan.	362.05	363.18	365.32	368.15	369.14	370.28	372.43	374.68	376.78	378.34	381.37	382.49	385.02	386.71	388.56
Feb.	363.25	364.00	366.15	368.87	369.46	371.50	373.09	375.63	377.36	379.66	382.02	383.72	385.87	387.17	389.98
Mar.	364.02	364.57	367.31	369.59	370.52	372.12	373.52	376.11	378.40	380.38	382.63	384.33	385.85	388.62	390.99
Apr.	364.72	366.35	368.61	371.14	371.66	372.87	374.86	377.65	380.50	382.14	384.40	386.23	386.77	389.51	392.52
May	365.41	366.80	369.30	371.00	371.82	374.02	375.55	378.35	380.61	382.24	384.94	386.43	388.49	390.18	393.22
Jun.	364.97	365.62	368.87	370.35	371.70	373.30	375.40	378.13	379.55	382.10	384.08	385.87	387.92	389.60	392.29
Jul.	363.65	364.47	367.64	369.27	370.12	371.62	374.02	376.61	377.77	380.66	382.37	384.44	386.32	388.01	390.49
Aug.	361.48	362.51	365.78	366.93	368.12	369.55	371.49	374.49	375.84	378.68	380.48	381.84	384.17	386.07	388.55
Sep.	359.45	360.19	363.90	364.63	366.62	367.96	370.71	372.98	374.05	376.40	378.78	380.86	383.00	384.61	386.54
Oct.	359.60	360.77	364.23	365.13	366.73	368.09	370.25	373.00	374.22	376.79	379.07	380.88	382.81	384.34	386.21
Nov.	360.76	362.43	365.46	366.68	368.29	369.68	372.08	374.35	375.84	378.32	380.17	382.40	384.06	386.02	388.61
Dec.	362.33	364.28	366.97	368.00	369.53	371.24	373.78	375.69	377.44	380.02	381.67	383.72	385.15	387.36	389.83

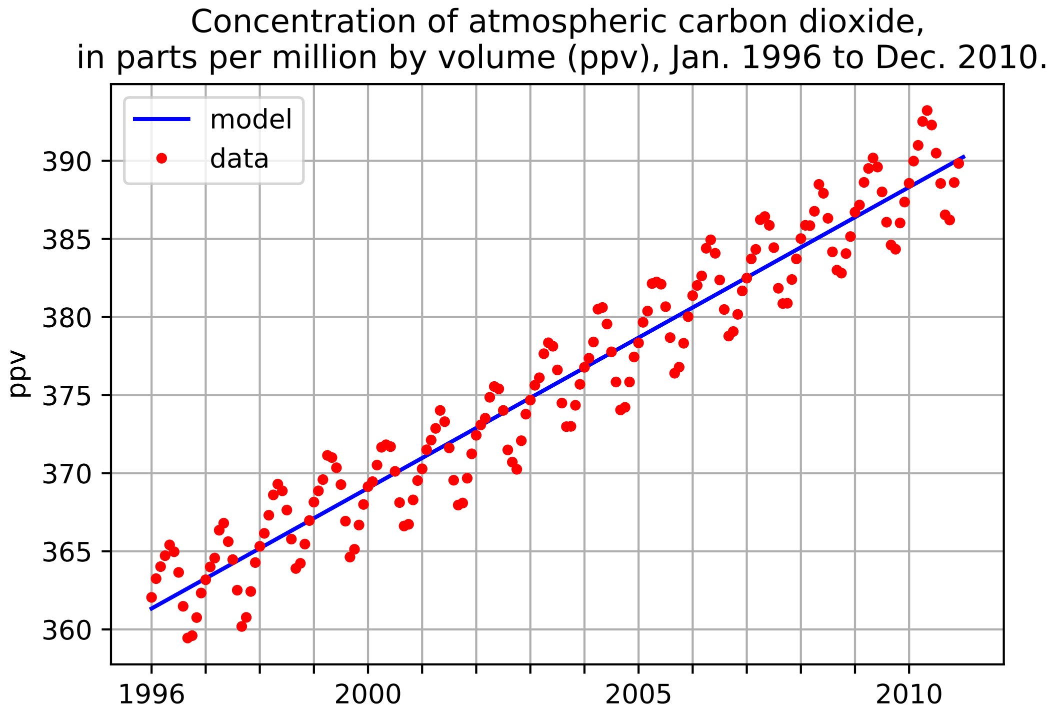
Concentration of atmospheric carbon dioxide, in parts per million by volume (ppv), recorded monthly from Jan. 1996 to Dec. 2010

Getting started: `https://colab.research.google.com/drive/1juTBIlQWf06NjXEzScfy2CNc4MPUJOS-`

Colab Question: The jupyter notebook below does the best fit with a function of the type

$$c_0 + c_1 t$$

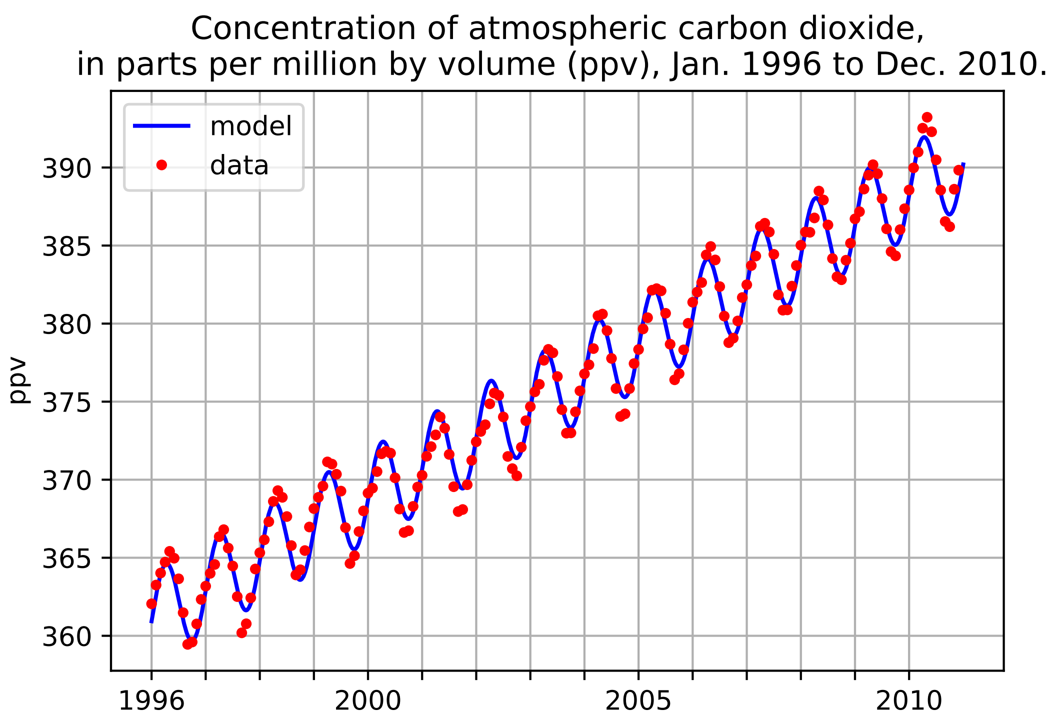
We get the following picture:



Change the model so that we get a best fit with a function of the type

$$c_0 + c_1 t + c_2 \cos(2\pi t) + c_3 \sin(2\pi t)$$

You should get the following picture:



You only need to change two lines of code:

```
A = np.array([ np.ones( t.shape ), t ]).T
```

and

```
yyy = 279. + x[0] + x[1] * xxx
```

Handwritten Question: Give the RMSEs of the two models $c_0 + c_1 t$ and of $c_0 + c_1 t + c_2 \cos(2\pi t) + c_3 \sin(2\pi t)$. Compare the RMSEs. Which model is better? Why?

EX.4.3.2.b, Sauer3

Apply Classical Gram-Schmidt orthogonalization to find the full QR factorization of the following matrix

$$\begin{bmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{bmatrix}.$$

EX.4.3.7.b, Sauer3

Use the QR factorization from Exercise EX.4.3.2.a, EX.4.3.4.a, or EX.4.3.6.a to solve the following least squares problem

$$\begin{bmatrix} -4 & -4 \\ -2 & 7 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 0 \end{bmatrix}.$$

EX.4.3.8.b, Sauer3

Find the QR factorization and use it to solve the following least squares problem

$$\begin{bmatrix} 2 & 4 \\ 0 & -1 \\ 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 2 \\ 1 \end{bmatrix}.$$

Suggestion: Use the Gram-Schmidt algorithm for getting the QR factorization, and remember that a thin QR factorization is enough for this problem. No need to compute a full QR factorization.

CP.4.3.6, Sauer3

Use **numpy** QR factorization to find the least squares solution and 2-norm error of the following inconsistent system:

$$\text{(a)} \begin{bmatrix} 3 & -1 & 2 \\ 4 & 1 & 0 \\ -3 & 2 & 1 \\ 1 & 1 & 5 \\ -2 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ -5 \\ 15 \\ 0 \end{bmatrix}, \qquad \text{(b)} \begin{bmatrix} 4 & 2 & 3 & 0 \\ -2 & 3 & -1 & 1 \\ 1 & 3 & -4 & 2 \\ 1 & 0 & 1 & -1 \\ 3 & 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \\ 2 \\ 0 \\ 5 \end{bmatrix}.$$

CP.4.5.1.b, Sauer3

Apply the Gauss–Newton Method to find the point $(\bar{x}, \bar{y})$ for which the sum of the squared distance to the three circles below is minimized. Use initial vector $(x_0, y_0) = (0, 0)$.

$$\text{Circles with centers } (-1, 0),\ (1, 1),\ (1, -1) \text{ and all radii } 1.$$

CP.4.5.2.b, Sauer3

Apply the Multivariate Newton's Method to the system (4.35) for the three circles below. Use initial vector $(x_0, y_0, K_0) = (0, 0, 0)$.

Circles with centers $(-1, 0)$, $(1, 1)$, $(1, -1)$ and all radii 1.

CP.4.5.3.b, Sauer3

Find the point (x, y) and distance K that minimizes the sum of squares distance to the circles with radii increased by K, as in Example 4.23.

Circles with centers $(-2, 0)$, $(3, 0)$, $(0, 2)$, $(0, -2)$ and all radii 1.

CP.4.5.4.b, Sauer3

Find the point (x, y) and distance K that minimizes the sum of squares distance to the circles with radii increased by K, as in Example 4.23. Plot the results.

$$\text{Circles with centers } (1,1),\ (1,-1),\ (-1,1),\ (-1,-1), \text{and } (2,0) \text{ and all radii } 1.$$

CP.4.5.8, Sauer3

Use the Levenberg-Marquardt Method with $\lambda = 1$ to fit a blood concentration model to the data of Example 4.11 without linearization.

CP.4.5.11.b, Sauer3

Apply Levenberg-Marquardt to fit the model $y = c_1 e^{-c_2 t}\cos(c_3 t + c_4)$ to the following data points, with an appropriate initial guess. State the initial guess, the regularization parameter λ used, and the RMSE. Plot the best least squares curve and the data points. This problem has multiple solutions with the same RMSE, since c_4 is only determined modulo 2π.

$$(t_i, y_i) = \{(1,2),\ (3,6),\ (4,4),\ (5,2),\ (6,-1),\ (8,-3)\}.$$