Task 1

Recall that you calculated sums in Task 1 in Session 1. Extend the code used for those sums to calculate the following.

Let $x_i = i$ and let $y_i = 2i + 0.3$ for $i = 1, 2, \ldots, 10$

- Calculate $\sum_i x_i$.
- Calculate $\sum_i y_i$.
- Calculate $\sum_i x_i y_i$.
- Calculate $\bar{x}$ and $\bar{y}$.
- Find the parameters a_0 and a_1 for a linear regression model of this data.
- Check your results are correct (you can use `matplotlib` library or Excel to plot the data and regression line together).

Task 2

When fitting a quadratic, the individual error between a data point and the estimated value is e_i as seen in

$$\begin{aligned} y_i &= a_0 + a_1 x_i + a_2 x_i^2 + e_i \\ \Rightarrow e_i &= y_i - a_0 + a_1 x_i + a_2 x_i^2 \end{aligned}$$

Let $f_i = a_0 + a_1 x_i + a_2 x_i^2$. Then the overall error function S can be written as

$$\begin{aligned} S &= \sum_{i=0}^{N-1} e_i^2 \\ &= \sum_{i=0}^{N-1} \left(y_i - a_0 + a_1 x_i + a_2 x_i^2\right)^2 \\ &= \sum_{i=0}^{N-1} (y_i - f_i)^2 \end{aligned}$$

To find optimal parameters, we need the partial derivative of S to be zero with respect to each parameter a_k. For example,

$$\frac{\partial f_i}{\partial a_0} = 1.$$

i) Find all partial derivatives $\dfrac{\partial f_i}{\partial a_k}$.

ii) Complete the following derivation for $\dfrac{\partial S}{\partial a_k}$ for fitting a quadratic.

$$\begin{aligned}\frac{\partial S}{\partial a_k} &= \frac{\partial}{\partial a_k}\sum_{i=0}^{N-1}\left(y_i^2 - 2y_i f_i + f_i^2\right)\\ &= \sum_{i=0}^{N-1}\frac{\partial}{\partial a_k}\left(y_i^2 - 2y_i f_i + f_i^2\right)\\ &= \dots\end{aligned}$$

Once you have a simplified sum, set the sum equal to zero, split up the sum into its components, and then form a general equation.

iii) Write out the set of three equations formed by letting $k = 1, 2, 3$.

iv) Transform the system of linear equations (in a_0, a_1, and a_2) into a matrix multiplication.

v) Using the (x_i, y_i) data on Blackboard and your Gaussian elimination program, solve the matrix multiplication $\mathbf{Ax} = \mathbf{b}$ you derived previously to fit a quadratic curve to the data.

Task 3

Instead of a function being in terms of powers of a single variable, the model may involve a linear combination of multiple variables $x_1, \dots, x_n$.

$$y_i = a_0 + a_1 x_{1,i} + a_2 x_{2,i} + \dots + a_n x_{n,i}$$

i) Similarly to before, continue the derivation of the set of equations when $y_i = a_0 + a_1 x_{1,i} + a_2 x_{2,i} + e_i$.

$$\begin{aligned} S &= \sum_{i=0}^{N-1} e_i^2\\ &= \sum_{i=0}^{N-1}\left(y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i}\right)^2\\ \Rightarrow \frac{\partial S}{\partial a_0} &= \dots, \frac{\partial S}{\partial a_1} = \dots, \frac{\partial S}{\partial a_2} = \dots.\end{aligned}$$

Use your Gaussian elimination program to solve the matrix multiplication $\mathbf{Ax} = \mathbf{b}$ formed by equating the partial derivatives to zero.

ii) Use the following data to solve for a_0, a_1, and a_2.

x_1	x_2	y
0	0	5
2	1	10
2.5	2	9
1	3	0
4	6	3
7	2	27

iii) Use `matplotlib.pyplot` to plot your function against the data to check the fit.

Task 4

Multiple linear regression is not limited to obviously linear data. How would you apply multiple linear regression to data you assume follows the relation $y = a_0 x_1^{a_1} x_2^{a_2} \dots x_n^{a_n}$?

Reading

You should (hopefully still) be reading additional material to provide a solid background as to the methods covered. Before next week, please read the introduction to part 5 and chapter 17 from Capra and Canale.

There is also lots of detail in chapters 14 and 15 of Numerical Recipes.