1. Given $n$ data points, $(x_1, y_1), \ldots, (x_n, y_n)$, the linear least-squares line of best fit is the linear function defined as,

$$f(x) = w_1 x + w_0$$

We now consider the sum of squares,

$$E(w_0, w_1) = \sum_{j=1}^{n} (y_j - f(x_j))^2$$

Expanding using the above definition of $f$ and squaring, we arrive at,

$$E(w_0, w_1) = \sum_{i=1}^{n} (y_i - w_1 x_i - w_0)^2 = \sum_{i=1}^{n} y_i^2 - 2w_1 \sum_{i=1}^{n} x_i y_i - 2w_0 \sum_{i=1}^{n} y_i + w_1^2 \sum_{i=1}^{n} x_i^2 + 2w_0 w_1 \sum_{i=1}^{n} x_i + w_0^2 \sum_{i=1}^{n} 1$$

In order to find the $w_0,\ w_1$ that minimize this function, we shall derive this equation with respect to both, and set equal to zero in order to calculate the critical points. First,

$$\begin{aligned}
\frac{\partial E}{\partial w_0} &= -2\sum_{i=1}^{n} y_i + 2w_1 \sum_{i=1}^{n} x_i + 2nw_0 && \text{Setting Equal to Zero and Solving} \\
0 &= -2\sum_{i=1}^{n} y_i + 2w_1 \sum_{i=1}^{n} x_i + 2nw_0 \\
2\sum_{i=1}^{n} y_i &= 2w_1 \sum_{i=1}^{n} x_i + 2nw_0 && \text{Divide by 2 and multiply by ``1''} \\
\frac{n\sum_{i=1}^{n} y_i}{n} &= \frac{nw_1 \sum_{i=1}^{n} x_i}{n} + nw_0 \\
n\bar{y} &= nw_1\bar{x} + nw_0
\end{aligned}$$

As desired. Next, we derive with respect to $w_1$,

$$\begin{aligned}
\frac{\partial E}{\partial w_1} &= -2\sum_{i=1}^{n} x_i y_i + 2w_1 \sum_{i=1}^{n} x_i^2 + 2w_0 \sum_{i=1}^{n} x_i && \text{Setting Equal to Zero and Solving} \\
0 &= -2\sum_{i=1}^{n} x_i y_i + 2w_1 \sum_{i=1}^{n} x_i^2 + 2w_0 \sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} x_i y_i &= w_1 \sum_{i=1}^{n} x_i^2 + w_0 \sum_{i=1}^{n} x_i && \text{Multiply by ``1'' as before} \\
\sum_{i=1}^{n} x_i y_i &= w_1 \sum_{i=1}^{n} x_i^2 + nw_0\bar{x}
\end{aligned}$$

As desired. Thus, we have found the minimizing equations. Next, we shall use Cramer's rule for $2 \times 2$ systems to solve for $w_0,\ w_1$. First, we place the above system into matrix form as,

$$\begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

Then the determinant of the coefficient matrix is,

$$D = n\sum_{i=1}^{n} x_i^2 - n^2\bar{x}^2$$

Substituting by Cramer's rule,

$$\begin{bmatrix} n\bar{y} & n\bar{x} \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

Then the determinant of the augmented matrix is,

$$D_{w_0} = n\bar{y}\sum_{i=1}^{n} x_i^2 - n\bar{x}\sum_{i=1}^{n} x_i y_i$$

And similarly,

$$\begin{bmatrix} n & n\bar{y} \\ n\bar{x} & \sum_{i=1}^{n} x_i y_i \end{bmatrix}$$

Then,

$$D_{w_1} = n\sum_{i=1}^{n} x_i y_i - n^2\bar{x}\bar{y}$$

Finally, we may compute,

$$\begin{aligned} w_0 &= \frac{D_{w_0}}{D} = \frac{n\bar{y}\sum_{i=1}^{n} x_i^2 - n\bar{x}\sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - n^2\bar{x}^2} \\ w_1 &= \frac{D_{w_1}}{D} = \frac{n\sum_{i=1}^{n} x_i y_i - n^2\bar{x}\bar{y}}{n\sum_{i=1}^{n} x_i^2 - n^2\bar{x}^2} \end{aligned}$$

2. We now consider the three distinct points, $(x_0, y_0),\ (x_1, y_1),\ (x_2, y_2)$. Given these points, we may construct a polynomial of degree 2 that will fit the data perfectly. As such, we know that for our points,

$$a_2x^2 + a_1x + a_0 = y$$

Will always be true. We now wish to write this system as a matrix equation. For our data points, we have the following equations,

$$\begin{aligned} a_2x_0^2 + a_1x_0 + a_0 &= y_0 \\ a_2x_1^2 + a_1x_1 + a_0 &= y_1 \\ a_2x_2^2 + a_1x_2 + a_0 &= y_2 \end{aligned}$$

Here, we note that the first entry in each equation have the common coefficient $a_2$ and so on across the left hand side of the equations. Thus, we see that this left hand side resembles a matrix equation with the entries for each equation being multiplied by an $a$ matrix that has the corresponding coefficeints. Thus, we shall write this system in the following form,

$$\underbrace{\begin{bmatrix} x_0^2 & x_0 & 1 \\ x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \end{bmatrix}}_{\mathbf{X}-matrix} \underbrace{\begin{bmatrix} a_2 \\ a_1 \\ a_0 \end{bmatrix}}_{\mathbf{a}-vector} = \underbrace{\begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}}_{\mathbf{y}-vector}$$

Or, in simpler form $\mathbf{X}\vec{a} = \vec{y}$, as desired.

3. We consider the potential equation,

$$1782^{12} + 1841^{12} = 1922^{12}$$

In SageMathCell, we compute the equation,

$$\sqrt[12]{1782^{12} + 1841^{12}}$$

To 11 digits of precision. The result,

$$\sqrt[12]{1782^{12} + 1841^{12}} = 1922.0000000$$

Based on this 11 digit precision, it would appear that the Theorem is incorrect.
We now use SageMathCell to compute the absolute error to 12 digits of precision,

$$|\sqrt[12]{1782^{12} + 1841^{12}} - 1922| = 4.37721610069e - 8$$

So, we see that at a slightly higher precision, the absolute error is a very small value. We may also compute,

$$\frac{|\sqrt[12]{1782^{12} + 1841^{12}} - 1922|}{1922} = 2.27742773189e - 11$$

Based on these error calculations, we see that this triple is very close to a true equation. In fact, on a lower precision integer format, rounding error would make it appear to be true. Thus we see that this may be called a Fermat near miss.

4. Completed

5. Completed