The Optimization of Linear Regression Through Quantum Algorithms

1. Introduction

Throughout many disciplines, it is important to obtain models that accurately depict the relationship between several variables. One common method to do so is known as regression analysis. This is a process that requires a general set of data. Using this set of data, a dependent variable is modeled as a combination of multiple independent variables (Gallo). Then, the adequacy of the model can be statistically verified through a series of rigorous tests. Afterwards, the model can be used to predict values or describe the relationship between an independent variable and a dependent variable. In addition, there are multiple types of regression analysis such as linear and non-linear, but for the scope of this paper, the focus will be on the former rather than the latter. For this process, there are numerous classical methods to compute these linear models that are readily available through statistical software such as R and Minitab. However, it is possible to significantly increase the computation speed of the models used in linear regression by utilizing a quantum algorithm. Notably, this procedure is able to outperform classical computation by taking advantage of the unique characteristics of quantum mechanics, but because of these same mechanics, there are limitations to its practicality.

2. Linear Regression

The goal of linear regression is to establish an accurate model of a dependent variable through a linear combination of independent variables from a certain data set. This can be expressed by the following equation:

$$(1)\ y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \varepsilon$$

where y is the dependent variable, β_0 is the y-intercept, every other β-value is a coefficient, x_i are the independent variables, n is the number of independent variables, and ε is the random error. Note that this is a probabilistic model due to the error term, so the deterministic component can be written as

$$(2)\ E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

While the independent variables are classified as x_i, it is important to consider that they are not always first order quantitative variables. For instance, they could also be higher order terms, interaction terms (the multiplication of two or more independent variables), or qualitative variables represented through a piecewise function. For linear regression, the deterministic component is the line or curve of best fit found through estimating the β-values. This estimation is done through an analysis of the given data set. Any deviances from the deterministic model is captured by the error term shown in equation 1.

After the model is complete, there is a large number of statistical tests that can be done to check its adequacy. First, the test of model utility, a hypothesis test on all of the β-values, can be done to see if the model is statistically significant. To fulfill this criterion, at least one of the independent variables must noticeably contribute to the prediction of the dependent variable. Then, the R^2-value can be analyzed to determine how well the model fits the data. Additionally, the coefficiants of variation can be observed to see if the range of the predictions is too large for practical use. Although there are many more statistical tests, these are a few of the more common ones conducted during linear regression analysis.

Once the adequacy of the model has been verified, it can be used for several practical applications. In particular, the model can be used to predict new values of the dependent

variables based on specific values of the independent variables. This can be done through interpolation or extrapolation. The former is a prediction made from the model that uses x_i-values within the initial range of the data, making it fairly accurate. The latter is a similar prediction except the x_i-values are outside the initial range of the data which causes the accuracy to be uncertain. This is because there might be a significant trend or error outside of the data set that has not been incorporated into the model. Thus, the predictions from the linear regression model are limited by the initial data set. Additionally, the strength of the contribution of the independent variables for a specific dependent variable can be assessed by considering the β-values. These interpretations are important for fields such as economics, medicine, and engineering where researchers are attempting to establish trends for large sets of data and determine the significance of certain variables.

3. The Method of Least Squares

One of the most common methods to estimate the β-values is through the method of least squares. This is a method that chooses the β-values based on the equation that minimizes the sum of squared residuals (Weisstein). In linear regression, a residual is the value found by subtracting the value of a dependent variable from its predicted value. This is the value represented by the vertical distance from the values of y to the line of best fit as seen in the following graph

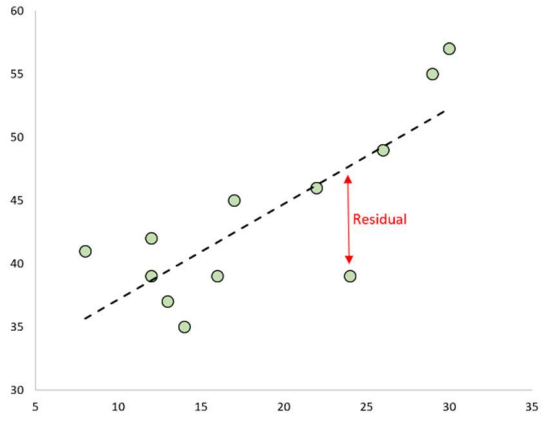


Fig. 1. Graph of residuals from: Zach.

Knowing this, the method of least squares can be mathematically expressed as

$$(3)\ \min(R) \text{ where } R = \sum_{i=1}^{n} (y_i - E(y_i))^2$$

In order to meet this requirement, the partial derivative of the squared sum of residuals must be 0 for every β-value. For the simple case of $E(y) = \beta_0 + \beta_1 x_1$, this can be seen by the following:

$$(4)\ \frac{\partial R}{\partial \beta_0} = -2\sum_{i=1}^{n}[y_i - E(y)] = 0 \text{ and } \frac{\partial R}{\partial \beta_1} = -2\sum_{i=1}^{n}[y_i - E(y)]x_i = 0$$

Thus, the β-values can be found by solving this system of equations for β_0 and β_1. While this is not difficult to reproduce for a small number of independent variables, the number of equations that need to be solved and their complexity rapidly increases with more independent variables, meaning the required algebra can quickly become overwhelming. Due to that, most researchers use statistical programs to compute these values such as R or Minitab.

In regard to these statistical programs, they also simplify this problem through the use of linear algebra a field of mathematics designed to solved systems of equations. This is done by first formatting the data as seen below where A is the matrix containing the values of the independent variables, b is the vector containing the dependent vales, and $\hat{x}$ is the vector that represents the weight of the independent variables. By solving for $\hat{x}$ as seen below, the method of least squares can be completed through the use of linear algebra which operates much quicker than solving for each equation individually (Rabinoff and Margalit).

$$(5)\ A\hat{x} = b \Rightarrow \hat{x} = (A^T A)^{-1} A^T b$$

Currently, the best known classical algorithm to solve for the linear system shown above runs in time $O(\sqrt{\kappa}sn)$ where κ is the condition number of A (the ratio between the largest and smallest numbers of A), s is the maximum number of nonzero entries in each row of A, and n is the number of independent variables (Liu and Zhang). While this is much better than other algorithms that run in exponential time, the time of the algorithm limits the maximum number of data that can be incorporated into any single model. Given a large enough data set, it would be impractical to use this method as the time it takes to sort through the data could take years.

For one final note on this method, it only considers the vertical distance between y and E(y) which allows for errors in the data from the dependent variable to disproportionally affect the model. One way to remedy this is to take the orthogonal distance from the values of the dependent variable to the line of best fit. This is a method called the total least squares which will not be discussed further within this paper as the focus is on the more common least squares method.

4. Quantum Computing and Its Benefits

To begin, quantum computing is the use of quantum properties inherent in subatomic particles in order to perform an algorithm at a speed than can sometimes be much faster than its classical counterpart. For these algorithms, quantum computers—devices capable of quantum computing—act on qubits using the quantum properties of superposition and entanglement. First off, a qubit is short for a quantum bit which is analogous to the bits used in classical computing. These are the 1s and 0s in computers that form into long strings of binary which are the backbone of any program. While classical bits are specific electrical pulses on a device, quantum bits are “subatomic particles such as electrons or photons” that are maintained in

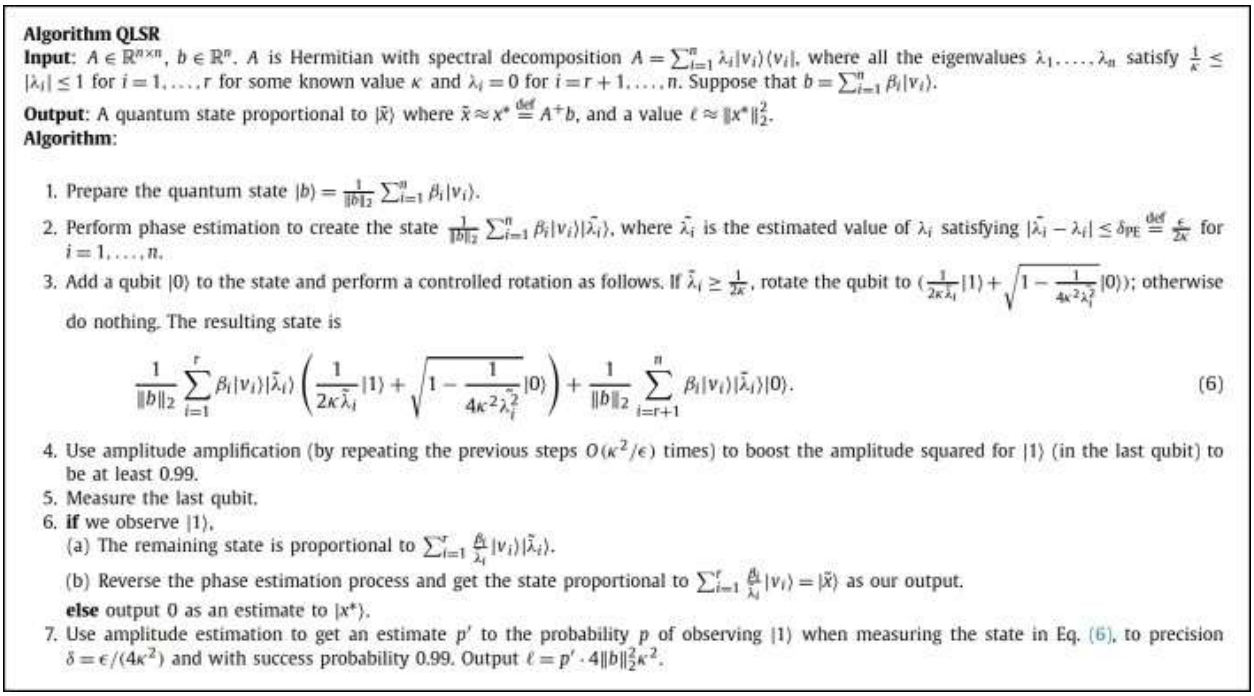
"superconducting circuits cooled to temperatures colder than deep space" or "electromagnetic fields on a silicon chip in ultra-high-vacuum chambers" (Giles). For the former quantum property, superposition is the ability for qubits to represent different combinations of 1s and 0s at the same time which allows for immense parallel processing. In this process, the qubits eventually consolidate once the superposition collapses to either 1 or 0. For the latter quantum property, entanglement is when two qubits are paired together to form a single quantum state where the manipulation of one qubit will predictably affect the other. For classical bits, the addition of a bit doubles the processing power, but for qubits, each additional qubit exponential increases its processing power due to entanglement (Giles). Thus, quantum algorithms can outperform classical algorithms in specific circumstances using the ability of qubits to undergo superposition and entanglement.

5. Quantum Algorithm for the Method of Least Squares

For quantum mechanics, one of the most common and well-defined ways to interpret it is through the Copenhagen interpretation. Because of this understanding of quantum mechanics, it is viable to mathematically interpret quantum algorithms as a linear combination of matrices. Knowing that, it is not to difficult to format the problem of the method of least squares in a way that is easily solvable for a quantum algorithm. This is because the method of least squares can be reduced to solving a matrix equation as shown above in section 3. However, it should be noted that the quantum algorithm outputs the solution "into a log n-qubit quantum state" rather than a single vector which has caused some controversy on the effectiveness of the algorithm (Liu and Zhang).

Before the algorithm runs, the matrix A is checked to see if it is Hermitian, meaning if it is equivalent to its conjugate transpose. If it is not, it is made Hermitian by inserting it into a

block matrix on the anti-diagonals while the diagonal entries are zero matrices. Subsequently, the b vector is appropriately changed. Then, the quantum algorithm is performed as stated in the graphic below:



Algorithm QLSR

Input: $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$. A is Hermitian with spectral decomposition $A = \sum_{i=1}^{n} \lambda_i |v_i\rangle\langle v_i|$, where all the eigenvalues $\lambda_1, \ldots, \lambda_n$ satisfy $\frac{1}{\kappa} \leq |\lambda_i| \leq 1$ for $i = 1, \ldots, r$ for some known value κ and $\lambda_i = 0$ for $i = r+1, \ldots, n$. Suppose that $b = \sum_{i=1}^{n} \beta_i |v_i\rangle$.

Output: A quantum state proportional to $|\tilde{x}\rangle$ where $\tilde{x} \approx x^* \stackrel{\text{def}}{=} A^+ b$, and a value $\ell \approx \|x^*\|_2^2$.

Algorithm:

1. Prepare the quantum state $|b\rangle = \frac{1}{\|b\|_2} \sum_{i=1}^{n} \beta_i |v_i\rangle$.
2. Perform phase estimation to create the state $\frac{1}{\|b\|_2} \sum_{i=1}^{n} \beta_i |v_i\rangle|\tilde{\lambda}_i\rangle$, where $\tilde{\lambda}_i$ is the estimated value of λ_i satisfying $|\tilde{\lambda}_i - \lambda_i| \leq \delta_{\text{PE}} \stackrel{\text{def}}{=} \frac{\epsilon}{2\kappa}$ for $i = 1, \ldots, n$.
3. Add a qubit $|0\rangle$ to the state and perform a controlled rotation as follows. If $\tilde{\lambda}_i \geq \frac{1}{2\kappa}$, rotate the qubit to $(\frac{1}{2\kappa\tilde{\lambda}_i}|1\rangle + \sqrt{1 - \frac{1}{4\kappa^2\tilde{\lambda}_i^2}}|0\rangle)$; otherwise do nothing. The resulting state is

$$\frac{1}{\|b\|_2} \sum_{i=1}^{r} \beta_i |v_i\rangle|\tilde{\lambda}_i\rangle \left(\frac{1}{2\kappa\tilde{\lambda}_i}|1\rangle + \sqrt{1 - \frac{1}{4\kappa^2\tilde{\lambda}_i^2}}|0\rangle \right) + \frac{1}{\|b\|_2} \sum_{i=r+1}^{n} \beta_i |v_i\rangle|\tilde{\lambda}_i\rangle|0\rangle. \quad (6)$$

4. Use amplitude amplification (by repeating the previous steps $O(\kappa^2/\epsilon)$ times) to boost the amplitude squared for $|1\rangle$ (in the last qubit) to be at least 0.99.
5. Measure the last qubit.
6. **if** we observe $|1\rangle$,
   (a) The remaining state is proportional to $\sum_{i=1}^{r} \frac{\beta_i}{\tilde{\lambda}_i} |v_i\rangle|\tilde{\lambda}_i\rangle$.
   (b) Reverse the phase estimation process and get the state proportional to $\sum_{i=1}^{r} \frac{\beta_i}{\tilde{\lambda}_i} |v_i\rangle = |\tilde{x}\rangle$ as our output.
   else output 0 as an estimate to $|x^*\rangle$.
7. Use amplitude estimation to get an estimate p' to the probability p of observing $|1\rangle$ when measuring the state in Eq. (6), to precision $\delta = \epsilon/(4\kappa^2)$ and with success probability 0.99. Output $\ell = p' \cdot 4\|b\|_2^2 \kappa^2$.

Fig. 2. Quantum algorithm for least squares from: Liu and Zhang.

To give a broad summary of the algorithm above, the first step prepares the quantum state. Then, the second step uses a phase estimation to tensor estimated eigenvectors to the previous quantum state. Afterwards, a 0 kat vector is added before a controlled rotation occurs which results in the state described above in the third step. This process is then repeated to amplify the amplitude of the desired expression until its probability of measurement is almost guaranteed. With that finished, the last qubit can be measured and used to estimate the compressed log-n qubit state that represents the weights for the method of least squares. Notably, this is done in time $O(log(n)s^2\kappa^3/\epsilon^2)$ where s and κ are defined as mentioned above while ϵ is the relative error.

6. The Issues with Quantum Computing

While there are many benefits in using a quantum algorithm for computation, there are also many issues. To begin, the quantum state is rather fragile. In particular, quantum states are extremely sensitive to their environment, so it can be difficult to manipulate an open quantum system. In addition, minor perturbances in the environment of a quantum system can cause it to collapse entirely, resulting in the system losing all of its quantum properties. This process is known as decoherence and needs to be considered for any quantum algorithm (Giles). Additionally, the theory of quantum algorithms is far past its physical counterpart. To be specific, many quantum algorithms have been verified through theoretical mathematics, but the building of an actual quantum computer has been exceedingly difficult. In fact, there are very few active quantum computers in the world, and of the ones that exist, they are unable to run algorithms that necessitate a high number of qubits. It should also be noted that quantum algorithms tend to be highly specialized as there are many simple calculations that can be done faster on a classical system. For instance, basic addition, subtraction, and multiplication can be done much faster through classical rather than quantum computation.

7. Conclusion

Linear regression is an important statistical tool to establish a relationship between a dependent variable and a series of independent variables using a set of data. One of the most common ways to perform it is though the method of least squares which estimates the β-values, resulting in a line or curve of best fit in time $O(\sqrt{\kappa}sn)$. Since this method involves solving a system of equations, it can be represented and solved rather simply through liner algebra. Once this model is determined, several statistical tests that can be used to check its adequacy. Afterwards, the model can be interpreted for predictions or to test the relationship between certain variables. Notably, the creation of the model can be done quicker through a quantum

algorithm that uses the properties of superposition and entanglement. Through the quantum algorithm outlined above, a model can be found in time $O(log(n)s^2\kappa^3/\epsilon^2)$. For large sets of data, this is an immense increase in speed. However, there are many issues with quantum computing such as the lack of working quantum computers and the difficulty in their setup. Due to this, the quantum algorithm for the method of least squares should be reserved for extremely large data sets when quantum computers become available in the future. While quantum computing provides many opportunities to speed up classical computing, the issues with its practical application force it to be limited for specialized cases.

Works Cited

Gallo, Amy. “A Refresher on Regression Analysis.” *Harvard Business Review*, 30 Nov. 2017, hbr.org/2015/11/a-refresher-on-regression-analysis.

Giles, Martin. “Explainer: What Is a Quantum Computer?” *MIT Technology Review*, MIT Technology Review, 2 Apr. 2020, www.technologyreview.com/2019/01/29/66141/what-is-quantum-computing.

Liu, Yang, and Shengyu Zhang. “Fast Quantum Algorithms for Least Squares Regression and Statistic Leverage Scores.” *Theoretical Computer Science*, vol. 657, 2 Jan. 2017, pp. 38–47., doi:10.1016/j.tcs.2016.05.044.

Rabinoff, Joseph, and Dan Margalit. “The Method of Least Squares.” *Interactive Linear Algebra*, textbooks.math.gatech.edu/ila/least-squares.html.

Weisstein, Eric W. "Least Squares Fitting." From *MathWorld*--A Wolfram Web Resource. https://mathworld.wolfram.com/LeastSquaresFitting.html.

Zach. “What Are Residuals in Statistics?” *Statology*, 7 Dec. 2020, www.statology.org/residuals.