

Name: Martha Geoffrey  
r-number: 0846297

## SCHEDULING

Planning is very important in my day-to-day life. I find it essential to structure my day, week, and month in advance to stay on track and ensure productivity. Normally, I would make a paper schedule with my school classes and other planned activities like weekly grocery shopping, room cleaning, appointments, etc., and I would also keep a to-do list of my assignments on the side. However, completing my assignments during my free time, i.e., when there are open slots on the schedule, proved to be ineffective because I would take longer than necessary to finish the assignments. Therefore, I reasoned that having a schedule that specifies what assignments must be completed when and for how long would be more beneficial than giving myself unlimited time to finish the assignments.

To determine the placement of my assignments in my weekly schedule, I would need the weekly schedule, training data, and new assignments' features. Assignments' features include assignment (name), priority, complexity, and workload. The features of previous assignments plus their duration minus their name make up the training data. The assignments in the training set are from the previous two months, and my estimation of their duration is fairly accurate. I used Multiple Regression to predict the duration of the new assignments which would be useful in the initialization stage of generating random schedules that includes the new assignments in Genetic algorithm. Since if the predicted duration of an assignment is three hours, it will have three slots in the schedule because one hour is represented by one slot in the schedules. I used Genetic algorithm to generate the fittest schedule, the schedule with the best arrangement of assignments according to my criteria from multiple possible solutions.

### Multiple Regression

Multiple linear regression was used to quantify the relationship between independent variables (priority, complexity, and workload) and dependent variable (duration) using training data. The following is a multiple linear regression model:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + e$$
$$y = XB + e$$

The following are steps by steps matrix approach to multiple linear regression that were performed to calculate the coefficients. The coefficients are used in predicting the duration of the new assignments. The formula to calculate the coefficients,  $\beta$ , is as follows:

$$\beta = (X^T X)^{-1} X^T Y$$

In the independent matrix, the columns represent features and rows represent observations, and a constant bias, 1, was added in the first columns of all rows. Then, the transposed independent matrix is multiplied by the independent matrix  $\Rightarrow (X^T X)$ . Then, the transposed independent matrix is multiplied by the array of dependent variable  $\Rightarrow (X^T Y)$ . The formula of calculating the inverse of  $(X^T X)$  is as follows:

$$(X^T X)^{-1} = \frac{1}{\text{Determinant of } (X^T X)} \times \text{Adjugate of } (X^T X)$$

The determinant of  $(X^T X)$  was first calculated, then the adjugated of  $(X^T X)$ , which equals to the cofactor of  $(X^T X)$ . Then, the coefficients are found as follows:

$$\beta = \frac{\text{Adjugate of } (X^T X) \times (X^T Y)}{\text{Determinant of } (X^T X)}$$

### Genetic Algorithm

To determine the placement of the new assignments in the schedule, a population of randomly generated solutions (chromosomes) or schedules was created. A fitness function was then defined to evaluate how well a schedule meets the requirements, and five best solutions were selected. Order crossover and mutation were then applied to evolve the population over a number of predetermined generations.

Name: Martha Geoffrey

r-number: 0846297

In defining a fitness function that evaluates how well a schedule (solution) meets the requirements. Three different approaches were used, adding up each approach's score to calculate the overall fitness of a schedule. First, if the assignments with a higher priority are scheduled for later days 'HighPrioritization' return -1. If a particular assignment is not scheduled in the schedule, 'AllAssignmentsIncluded' returns -1. Thirdly, if the predicted duration of a particular assignment differs from the actual predicted duration in the schedule, 'AccuratePosition' return -1.

To choose five best solutions, the list of 'ScheduleChromosome' objects is sorted using bubble sort in descending order according to their 'Fitness' property. The top five solutions are then selected. Because sorting the solutions, in my opinion, equates to ranking them, I view sorting the solutions as the rank selection's implementation. In ranking selection, the concept of a fitness value is removed, and "every individual in the population is ranked according to their fitness. The selection of the parents depends on the rank of each individual and not the fitness. The higher ranked individuals are preferred more than the lower ranked ones." The population's individuals have very close fitness values at the end of the run, so choosing the best five makes absolutely sense.

Order crossover is chosen to preserve the order of the assignments as it ensures certain patterns from the parents are retained in the offspring. Order crossover is done by selecting "a substring from one parent at random, produce a proto-child by copying the substring into the corresponding positions as they are in the parent, delete all the {assignments} from the second parent {that} are already in the substring, the resultant sequence contains the {assignments} the proto-child needs, {lastly} place the {assignments} into the unfixed positions of the proto-child from left to right according to the order of the sequence to produce an offspring."

The mutation is done by selecting two assignments with the same predicted duration at random from the offspring schedule and swap their positions. For example, if two distinct assignments have a predicted duration of three hours (slots) each, that means the positions of those slots will be swapped. Although, if the maximum number of attempts has been reached and there are no two distinct assignments with the same predicted duration, two slots containing distinct assignments will be chosen at random and their positions will be swapped.

## Challenges

Coming up with a fitness function was difficult because there isn't a fitness function that works for all problems; instead, it depends on the specifics of each one. So, I am not sure what I did is correct. To ensure that the fittest schedule contains all assignments, consecutive slots of the same assignment, and the number of slots of the same assignment equals the predicted duration proved to be challenging as well. This is mostly because during crossover, I took the first and second parts of the parent 1 and parent 2 schedules, which causes some assignments to be multiplied more than others. Additionally, during the mutation, I chose an assignment in the offspring schedule at random and replaced it with a random assignment from the list of assignments objects, which made my issue even worse. Fortunately, the sources mentioned taught me about order cross over and the now-implemented mutation. Another problem that arose due to order crossover was that if there are fewer assignments, one half of the week is fuller than the other half, while I would like the assignments to be distributed equally throughout the week.

## Ethical Aspects

Ethical aspects to keep in mind is that there might be potential biases in the training data used to training the regression model because the training data is not diverse, it's personal data from one person. There are also privacy concerns of the individual whose data is used in the scheduling process as it might be used incorrectly used in the wrong hands. Therefore, we must ensure that the data is handled with care and stored and processed securely. Also, there is lack of transparency in the training data since it's difficult to know if it's only school, personal, or work assignments. Additionally, we also don't know the accuracy of the regression model as it was not evaluated. In conclusion, I started noting the predicted and actual durations of the assignments to improve the model's performance.

Name: Martha Geoffrey  
r-number: 0846297

## References

- [https://www.youtube.com/watch?v=s\\_p3pdmyX8s&ab\\_channel=Stabelm](https://www.youtube.com/watch?v=s_p3pdmyX8s&ab_channel=Stabelm)
- [https://www.youtube.com/watch?v=GMvhcEs2dh4&ab\\_channel=BeginCodingFast](https://www.youtube.com/watch?v=GMvhcEs2dh4&ab_channel=BeginCodingFast)
- [https://www.youtube.com/watch?v=H9BWRyJNlv4&ab\\_channel=KhanAcademy](https://www.youtube.com/watch?v=H9BWRyJNlv4&ab_channel=KhanAcademy)
- <https://core.ac.uk/download/pdf/32452919.pdf>
- <https://taytzehao.medium.com/classroom-scheduling-using-heuristics-and-genetic-algorithm-b7279ecdd74c>
- <https://intapi.sciendo.com/pdf/10.2478/v10238-012-0039-2>
- <https://www.baeldung.com/cs/genetic-algorithms-roulette-selection>
- [https://www.tutorialspoint.com/genetic\\_algorithms/genetic\\_algorithms\\_parent\\_selection.htm](https://www.tutorialspoint.com/genetic_algorithms/genetic_algorithms_parent_selection.htm)
- <https://www.statology.org/multiple-linear-regression-by-hand/>
- <https://www.baeldung.com/cs/ga-order-one-crossover#:~:text=We%20use%20the%20order%20one,the%20chromosomes%20of%20two%20parents.>

End.