Introduction to Machine Learning

許木羽 / 111000177

1. **Basic Part**
   OLS Equation: $0.004611890733424811x^2 + 0.23531354225880863x$ 77.90204667432596
   Gradient Descent Equation: $0.005957476670683341x^2 + 5.9157888855214024e-05x$ 88.00000049050026

2. **Advance Part**
   In the basic part, we only use 1 input variable.
   In the advance part, we use 4 inputs (temperature, heartrate, resprate, o2sat). Additionally, we include some input from all k data before as input, for k as integer (in this case I set k to 5) as we make it become auto regression. Also, we split every patient as they have different gradient functions.

3. **Difficulty Encountered and Solutions**
   a. Gradient descent constant function isn't increasing significantly as we hope.
      Solution: Iterate every constant initializer, or make the learning rate for constant higher
   b. MAPE for the advanced part is not as good as expected.
      Solution: Use different learning rate for different variables
   c. Temperature has so many null values.
      Solution: We might drop it, but we need to analyze it first. My best approach is fill the null value with the same data that has closest chart time with the null value time. But it's complicated, so we can just fill all null values with average. Other than temperature we can drop as it doesn't consist too much null value, but still possible to fill the null value with closest data.
   d. Remove outliers.
      Solution: as I don't have any idea what SBP is, it would be bad if I falsely assume the upper and lower bound, so I remove all the outlier (excluding null value) using Z-value. Anything has bigger than 3 or smaller than -3 is considered out of range.

**Reflection:**

Data science fundamentals might improve your machine learning model as you need to analyze first your data. You need to carefully analyze the data and not randomly drop the data. OLS is good and easy, and gradient descent might be tricky, but rewarding. It means if you imply the linear descent good enough, it can search for a better solution than OLS.