May 20

**What I’ve worked on**: I found an interesting dataset on Kaggle that I planned to work on: customer\_experience\_data.csv. Also, I made my research question clear. My project will focus on building *machine learning algorithms* to (1) predict satisfaction scores based on the 4 numerical independent variables and (2) decide which independent variable is most important in terms of predicting the satisfaction score. I checked my thoughts with the TA in the tutorial. I have also started to build the multivariant linear regression model from scratch.

**What problems I encountered**: I only think about linear regression and polynomial regression, and I am not sure what other models I can use.

**What I learned**: (1) How to use Kaggle to find datasets (2) The overall structure of building linear regressing using gradient descent.

**Which resources did I use**: Kaggle, Google.

May 21

**What I’ve worked on**: I built a function to clean the data. I built a linear regression machine learning algorithm from scratch, in order to find the coefficients w and b. I started to research on multivariant polynomial regression (to the power of 2) from scratch.

**What problems I encountered**: I don’t know how to build the multivariant polynomial regression: I don’t know how to transform the X matrix (independent variable) in order to fit a polynomial regression, and I don’t know what the formula of the polynomial regression should be.

**What I learned**: (1)How to build a gradient descent linear regression from scratch. (2) The overall structure of building a polynomial regression. (3) How to fill NA values using the mean. (3) How to normalize the input data.

**Which resources did I use**: Google.

May 22

**What I’ve worked on**: I finished the linear regression model by implementing two functions: 1. Access prediction based on w and b. 2. Assess the performance of the model by using RMSE. I did research on how to access the performance of a regression model before working on these functions.

**What problems I encountered**: I was still not sure what other model I could use, and I did not know how to build multivariant polynomial regression from scratch.

**What I learned**: (1)How to assess the performance of a regression model. (2) How to build such an algorithm from scratch. (3) How to use the library “from sklearn.metrics import mean\_squared\_error”

**Which resources did I use**: Google.

May 23

**What I’ve worked on**: I checked the problems that I cannot solve with TA, and we came to a conclusion that: (1) Multivariant polynomial regression is mathematically too difficult to implement from scratch, but I could choose to use a library. (2) A decision tree is good for deciding which independent variable is most important. (3) Another model to fit the multivariant data is needed. I first focused on how to build a decision tree for regression from scratch. I started by understanding its working mechanism.

**What problems I encountered**: I understand the mechanism of decision tree in classification problems, but not in regression.

**What I learned**: (1)How does a decision tree work. (2) The overall structure of building a decision tree algorithm.

**Which resources did I use**: Google, YouTube (for the introduction of the decision tree algorithm).

https://www.youtube.com/watch?v=ZVR2Way4nwQ&list=PLM8wYQRetTxAIU0oOarQeW2WOeYV9LyuG&index=8

https://www.youtube.com/watch?v=P2ZB8c5Ha1Q

May 24

**What I’ve worked on**: I figured out how to build a decision tree regressor from scratch and finished the code. I finished comparing the independent variables on their importance. I finished implementing the multivariant polynomial regression, using the sklearn library.

**What problems I encountered**: In the learning video of building a decision tree from scratch, the data structure was different from my codes and the code example in the video. However, I have managed to write the codes that fit my data structure.

**What I learned**: (1)How to build a decision tree from scratch. (2) library: “from sklearn.preprocessing import PolynomialFeatures”, “from sklearn.linear\_model import LinearRegression”

**Which resources did I use**: Google, YouTube: https://www.youtube.com/watch?v=ZVR2Way4nwQ&list=PLM8wYQRetTxAIU0oOarQeW2WOeYV9LyuG&index=8

https://www.youtube.com/watch?v=P2ZB8c5Ha1Q

May 25

**What I’ve worked on**: I finished the last multivariant model: k-Nearest Neighbors (KNN) for regression: build from scratch. I compared all three multivariant models, to see which of them predicts the best. Also, I worked on the project description.

**What problems I encountered**: How to arrange the code to make it structural and understandable.

**What I learned**: How to build a KNN model for regression from scratch.

**Which resources did I use**: Google.

May 26

**What I’ve worked on**: I finished the project description, finalized all the project files, and prepared for the presentation.

**What problems I encountered**: NA

**What I learned**: Better understanding of my entire project.

**Which resources did I use**: NA

May 27

**What I’ve worked on**: I experimented with adding the train accuracy of all models in research question 1, and the results were surprising. I change the codes and project description accordingly.

**What problems I encountered**: NA

**What I learned**: Better understanding of the models.

**Which resources did I use**: NA