111062118 江佩霖

Machine Learning

Lab1 Report

I. Basic

A. Variables and the Regression equation

$$y = wx + b$$

$$(w, b) = (8.95505, 44.06432)$$

僅有一個特徵值(weight)和常數 b 的 equation

II. Advanced

A. Regression equation

$$y = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6 + w_7x_7 + b$$

7 個特徵值(分別對應 age, gender, height, weight, bodyfat, diastolic, systolic)和常數 b

B. Variables

- Male (w_m, b_m) :

  w = [-0.29431, 0.00000, -0.45185, 4.90928, -3.97597, 0.53900, 0.14364], b = 52.07306
- Female (w_f, b_f):

  w = [-0.89423, 0.00000, 0.03358, 2.91151, -2.36817, 0.24561, 0.30890], b = 30.92832

III. Conclusion

1. **Matrix Shape Mismatches:** As I am not very proficient in Python, I often encounter issues with mismatched shapes when handling matrix data. I frequently seek assistance from ChatGPT or friends to identify and correct these errors.
2. **Gradient Descent Issues:** I struggled to reduce the MAPE value using only gradient descent, possibly due to my lack of experience with higher-order gradient descent implementations. I eventually used the matrix computation methods discussed in class, which successfully reduced the MAPE in Basic.
3. **Advanced Problems:** Due to the numerous parameters involved, I faced persistent issues with the first problem and had to be careful when processing data while splitting modules. This often led to mistakes. I decided to set this aside for now and will revisit it when I can approach it with more patience.
4. **Data Cleaning:** I initially followed guidance to handle missing values by using mean imputation. However, after hearing from a friend that this could result in unappealing visualizations, I switched to removing NaN values and extreme values instead.

IV. Summarize

1. **Importance of Data Cleaning:** I discovered that effective data cleaning is crucial, as it can significantly reduce the MAPE value.
2. **Understanding Model Optimization:** I learned that tuning model parameters and choosing the right optimization techniques can greatly impact model performance. The process of adjusting learning rates and using advanced methods like Ridge regression has shown me the importance of fine-tuning and validating different aspects of the model to achieve better results.