# HW3 Programming

# 109061225 許博翔

## 1. Data Pre-process

In this part, we need to handle two problems. First is the conversion of non-Number type to Number type, the other is the construction of dataset.

The useful features contains 'Gender', 'Age', 'Height', 'Weight', 'Duration', 'Heart_Rate' and 'Body_Temp' while Gender is non-Number type. So, we need to use map function to turn male to 0 and female to 1.

To construct the dataset, we need to split the whole data into three parts, train, validation, and test with specific ratio, 7:1:2. After that, we need to split the Calories from each data as answer.

```
Mean Squared Error of MLR: 128.67755671507882
Mean Squared Error of BLR: 133.07754538497568
Mean Squared Error of Neural Networks: 6.470630
Mean Squared Error of SVR: 262.760096
Mean Squared Error of Ridge: 128.677418
Mean Squared Error of Elastic Net: 154.089339
Mean Squared Error of Linear Regression: 128.677557
Mean Squared Error of Degree-2 Polynomial Regression: 8.727518
Mean Squared Error of Degree-3 Polynomial Regression: 0.085154
Mean Squared Error of Degree-4 Polynomial Regression: 0.087938
Mean Squared Error of Decision Tree Regression: 31.769410
Mean Squared Error of Random Forest Regression: 8.493131
Mean Squared Error of Gradient Boosting Regression: 8.679973
Best model is Degree-3 Polynomial Regression with MSE 0.085154
```

Loss of each Regression Model

## 2. Maximum Likelihood Regression (MLR)

By the formula given in HW3.pdf, we implement it. The method I used is Ordinary Least Squares (OLS) and use the Mean Square Error (MSE) as the training loss.

We enhanced the input features by including a column of ones for the bias term to implement the MLR. The line that best fits the training data was then found using the normal equation, a closed-form solution to linear regression.

# 3. Bayesian Linear Regression (BLR)

Estimation methods used by the MLR and BLR are the main distinctions between them. The BLR provides a distribution over potential lines, whereas the MLR estimates a single best-fitting line. With the addition of a column of ones to the input features, the BLR model was implemented similarly to the MLR. The BLR, on the other hand, was unique because it made use of batch training, which made it possible to compute the posterior predictive distribution at each step and produce a distribution over potential regression line.

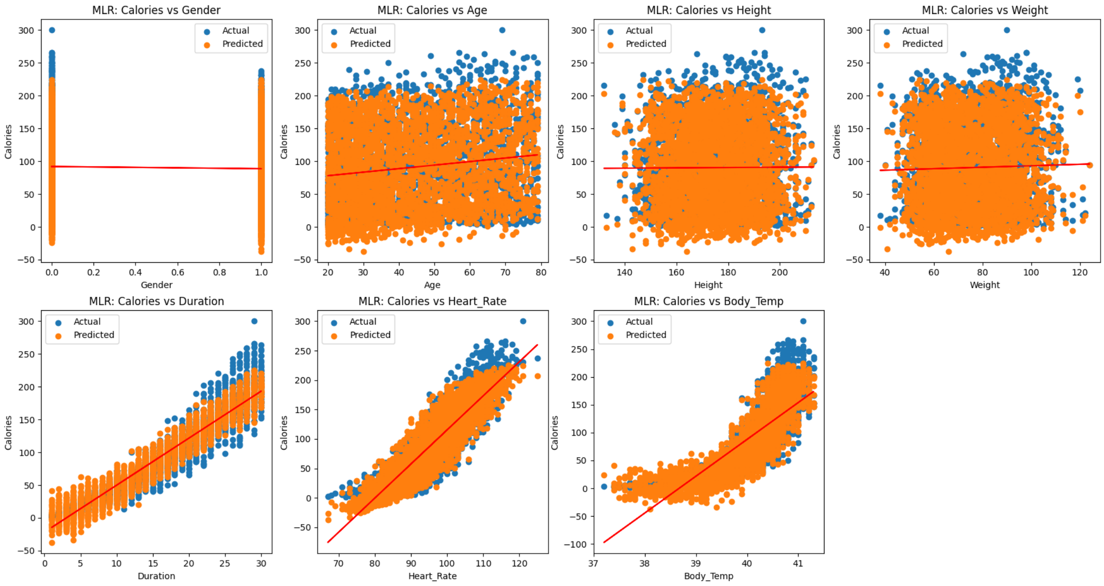
Same as above, we use MSE for training loss.

# 4. Discussion

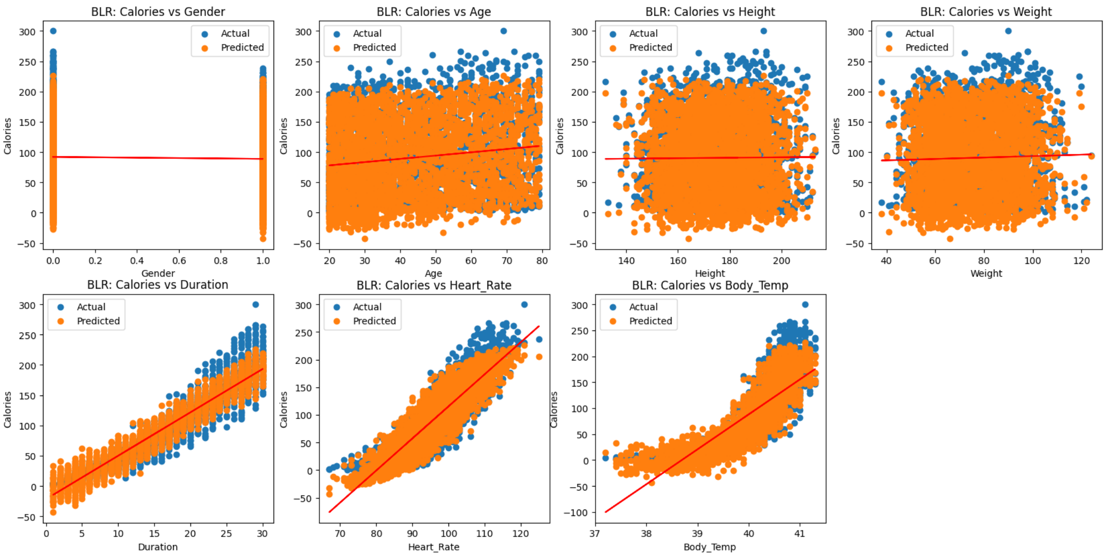
Both of MLR and BLR are statistical methods used for estimating the parameters of a model, but they approach the problem in different ways.

Maximum Likelihood Estimation (MLE) maximizes a likelihood function to estimate a model's parameters. The objective of linear regression is to identify the parameters (the independent variable coefficients) that maximize the likelihood of the observed data given the model. The MLE approach assumes that the parameter's estimated value is fixed but unknown. It does not consider prior assumptions about the parameters and, unless you compute a confidence interval, it only provides a point estimate without a measure of uncertainty.

Bayesian methods, on the other hand, treat the parameters as random variables. When performing Bayesian linear regression, we begin by creating a prior distribution that reflects our assumptions regarding the parameters prior to viewing the data. We revise our assumptions considering the data in order to obtain a posterior distribution of the parameters. The likelihood function and the information from the data (through the posterior distribution) are both incorporated. This method offers a complete distribution of potential parameter values rather than just a point estimate, enabling a more thorough understanding of uncertainty.




The best-fit line of MLR with each feature




The best-fit line of BLR with each feature

# 5. Other Regression Model

To find the best Regression Model, I train and use some regression models such as NN, Ridge or Liner Regression. I referenced some of the responses from the web and the generative model. Of all the models I have chosen, the best performers are Degree-3 Polynomial Regression Model with MSE 0.085154.

```
Mean Squared Error of MLR: 128.67755671507882
Mean Squared Error of BLR: 133.07754538497568
Mean Squared Error of Neural Networks: 6.470630
Mean Squared Error of SVR: 262.760096
Mean Squared Error of Ridge: 128.677418
Mean Squared Error of Elastic Net: 154.089339
Mean Squared Error of Linear Regression: 128.677557
Mean Squared Error of Degree-2 Polynomial Regression: 8.727518
Mean Squared Error of Degree-3 Polynomial Regression: 0.085154
Mean Squared Error of Degree-4 Polynomial Regression: 0.087938
Mean Squared Error of Decision Tree Regression: 31.769410
Mean Squared Error of Random Forest Regression: 8.493131
Mean Squared Error of Gradient Boosting Regression: 8.679973
Best model is Degree-3 Polynomial Regression with MSE 0.085154
```

The reason why this model will be the best-performance maybe due to those reason below:

(1) Non-linearity: Non-linear relationships are frequent in real-world data. Compared to a straightforward linear model, a degree-3 polynomial can represent more intricate relationships between the dependent and independent variables. It can detect data curvature that a linear model might not be able to.

(2) Overfitting and Underfitting. A model that is too simple, such as a linear one, may not be able to account for all the patterns in the data (underfitting), while a model that is too complex, such as a high-degree polynomial, may be able to account for noise and outliers but be less successful at generalizing to new data. A polynomial of degree 3 might offer a good compromise.

(3) Empirical Evidence: It's possible that a degree-3 polynomial was discovered to perform best on this particular dataset through exploratory data analysis and model testing. The specific distribution and relationships in the data may be to blame for this.

(4) Features that Interact and Combine: A degree-3 polynomial can account for interactions between various features up to a third degree. This is helpful if, for instance, the relationship between the effects of height and weight on calories or the relationship between duration and heart rate on calories changes.