Logistic Regression

Assignment 3  
Name: Zihan Yan

**Determine the data dimensionality**

* Total number of patients: 169
* Number of attributes (categories): 5
* Data types.
  + Age float64
  + Weight\_kg float64
  + Height\_cm float64
  + BMD float64
  + Fracture object
* Missing values.
  + Age 0
  + Weight\_kg 0
  + Height\_cm 0
  + BMD 0
  + Fracture 0
* Number of patients in each target class.
  + no fracture 119
  + fracture 50

# Apply Logistic Regression using Ridge Regulation

* Feature(s) considered important based on the coefficient values.
  + Feature importance indicated by coefficient:
    - Age -0.046444
      * it shows a small negative association between age and fracture risk.
    - Weight\_kg 0.066782
      * Although positive, the coefficient for body weight is considerably lower than for BMD, suggesting a modest contribution to fracture risk, with greater body weight marginally raising the likelihood of fractures.
    - Height\_cm -0.014642
      * The negative coefficient for height indicates a minor decrease in fracture probability with increasing height, yet an inverse relationship to fracture risk compared to BMD.
    - BMD 2.382925
    - BMD is a primary feature, with a high positive coefficient indicating a strong correlation.
* Classification accuracy.
  + 0.8431372549019608
* Number of patients misclassified for each target class.
  + fracture 6
  + no fracture 2

# Apply Logistic Regression using Lasso Regulation

* Feature(s) considered important based on the coefficient values.
  + Age -0.034442
    - Slightly reduces fracture risk. It suggests a minor protective effect as age increases.
  + Weight\_kg 0.041916
    - Marginally increases fracture risk. It has a much smaller effect than BMD.
  + Height\_cm -0.032225
    - Shows a minimal decrease in fracture risk with increasing height.
  + BMD 7.633766
    - a very high positive correlation between higher BMD values and increased fracture likelihood
* Classification accuracy.
  + 0.8627450980392157
* Number of patients misclassified for each target class.
  + Misclassified
  + fracture 5
  + no fracture 2

**Comparison of classification accuracies among the regulation methods.**

* The accuracy of Ridge: 0.8431372549019608
* The accuracy of Lasso: 0.8627450980392157
* Lasso regularization achieves a superior classification accuracy of 86.27% compared to Ridge's 84.31%, likely benefiting from its feature selection capability, which simplifies the model by zeroing out less relevant coefficients.