The MICHD Classification Effort

Aurelien Ferlay

Autumn 2023, EPFL, Switzerland

Summary

For a mandatory segment of the Machine Learning curriculum at EPFL, we utilized machine learning approaches on actual medical statistics procured by BRFFS. Our objective was predicting individuals with coronary heart disease (MICHD). This document elaborates on our analytical results, the techniques we employed, and our handling of the expansive initial dataset.

1 Preamble

Cardiovascular ailments, as documented by the World Health Organization, encompassing stroke and heart attack instances, have now become a leading cause of mortality worldwide. With rising average age, heart-related disorders are growing in prevalence. Machine learning can aid in early detection and potentially warding off cardiovascular disease onset. Our project's primary purpose was to evaluate an individual's likelihood of developing cardiovascular issues by analyzing lifestyle indicators. Post dataset comprehension and refinement, we utilized five distinct machine learning algorithms for predictions. This document showcases the performance metrics of all models, pinpointing the most effective model and its specific configurations. We then conclude with our final insights.

2 Techniques Employed

Five main functions were coded in the implementations.py file :

- Gradient Descent
- Stochastic Gradient Descent

- Ridge Regression
- Logistic Regression (through gradient descent for Log Likelihood maximization)
- Regularized Logistic Regression