

SAN: Review of Team G

Adrian Filčík, Adam Hrabovský, Zuzana Világiová, Matej Vojtek

January 2024

1 Summary

The study aimed to develop a method for predicting the jugulum-umbilicus and jugulum-symphysis distances using specific previously available and precisely measured patient data, without the need to manually measure the distances using a tape measure. The primary objective was to reduce errors associated with manual measurements, thereby enhancing the accuracy of estimating pulse wave propagation speed. This improvement could contribute to a more precise diagnosis of the stiffness of the patient's vascular wall.

The methodology involved data preprocessing, exploring data modality, reweighing to address high variance, and building a basic Generalized Linear Model (GLM) with fundamental predictors. The team extended the model by incorporating additional predictors, applied regularization with LASSO to select important features, and conducted backward stepwise selection for sequential feature selection. Models were evaluated based on Mean Squared Error (MSE) and Mean Absolute Error (MAE), and the results were compared to determine the most effective approach. The discussion highlighted the feasibility of accurate predictions without measuring specific distances and identified key predictors such as height, weight, BMI, and arm circumference. The references cited relevant literature in the context of cardiovascular diseases and arterial stiffness, providing a comprehensive solution to the assignment.

The SAN Final Assignment reflects a commendable collaborative effort, showcasing a systematic approach to cardiovascular health prediction. The team's adept handling of data thorough model evaluation and demonstrate strong analytical skills.

2 Remarks

- Can you elaborate more on the reweighing in Section 3.3? Why do you assume it will help? How exactly did you arrive at the presented value?
- How exactly does the best model you found compare to the current commonly-used manual measurements? The MAE is less than 3 cm, but is this actually useful?
- How are the anatomical distances in the dataset measured? Are they reliable enough to use for the model?
- In the Section 5.2 you used cross-validation to find the optimal α , did you also use cross-validation to test the other models?

3 Other details

- Why say GLM if you're just using linear regression?
- It would probably be more clear if the predictor coefficients were given in a table rather than writing out the equation