Phase 4

Feature Importance and Reduction

Alexander Garcia

December 2024

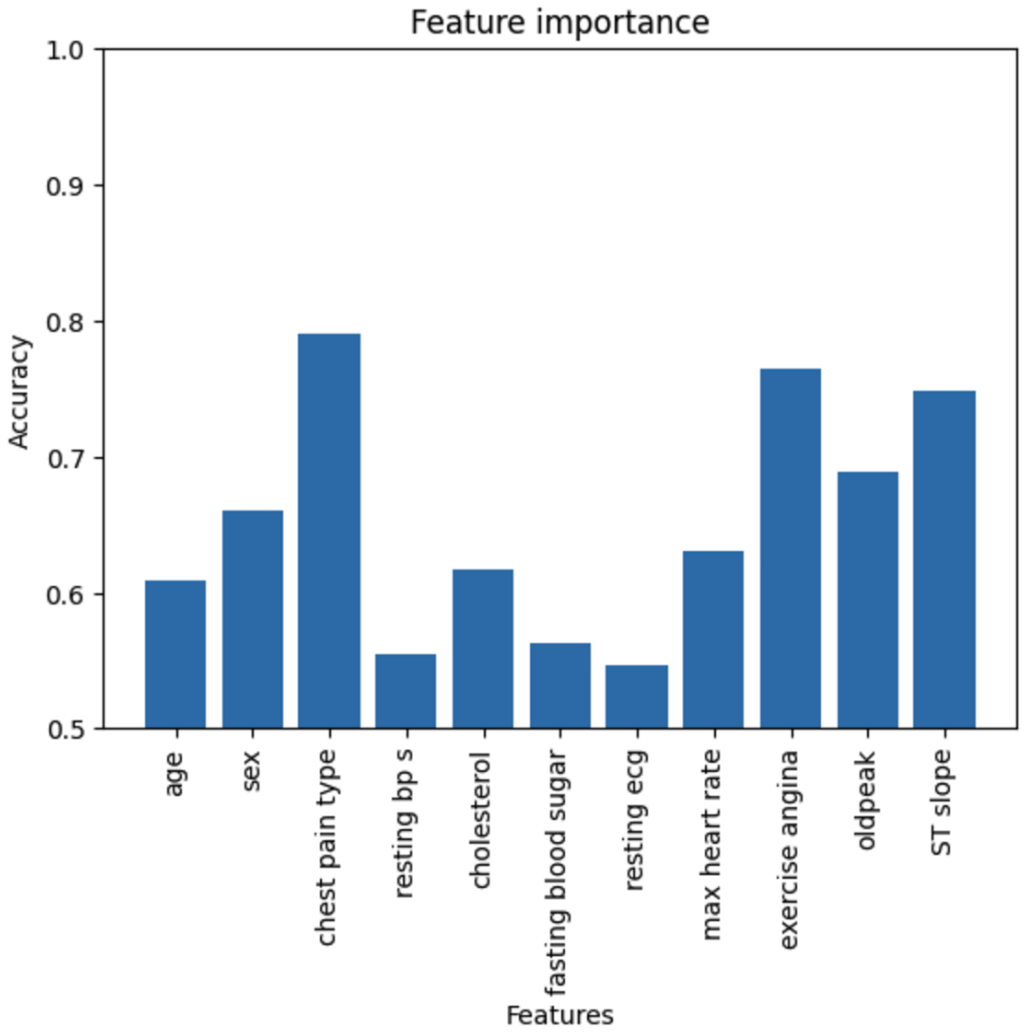
Contents

1 Introduction

The goal of this phase is to find feature importance by running models with just one feature. Meaning, for my 11 feature data set I want to run 11 different models testing one feature at a time. Also, feature reduction, is then iteratively removing the features one at a time to see if there is impact on accuracy. This will help determine which features provide the biggest impact to training and which features could possibly be removed. The model chosen to run feature importance and reduction with was the 8-8-8-1 model.

2 Feature Importance

Looking at the graph below, there are several features that are not contributing much to the accuracy of the model. Resting blood pressure, fasting blood sugar, and resting ECG results are all less than 60% accuracy when run alone. Meanwhile, 3 features seem to have a significant impact on the final results measuring above 70% accuracy when run alone.



3 Feature Reduction

The process involves iteratively removing features. To start, I removed features that did had lower accuracies when run in the feature importance step. I did not expect an increase in accuracy here and there was a slight decrease. Removing cholesterol did increase accuracy slightly. The biggest surprise was in the removal of features which had higher accuracy in the feature importance phase. I did expect bigger drops in accuracy when removing a feature which scored above 70% accuracy when run in its own model. However, sig-

nificant drops to accuracy did not occur until removing 8 of 11 features which could mean the model is overfitting and learning the pattern too closely.

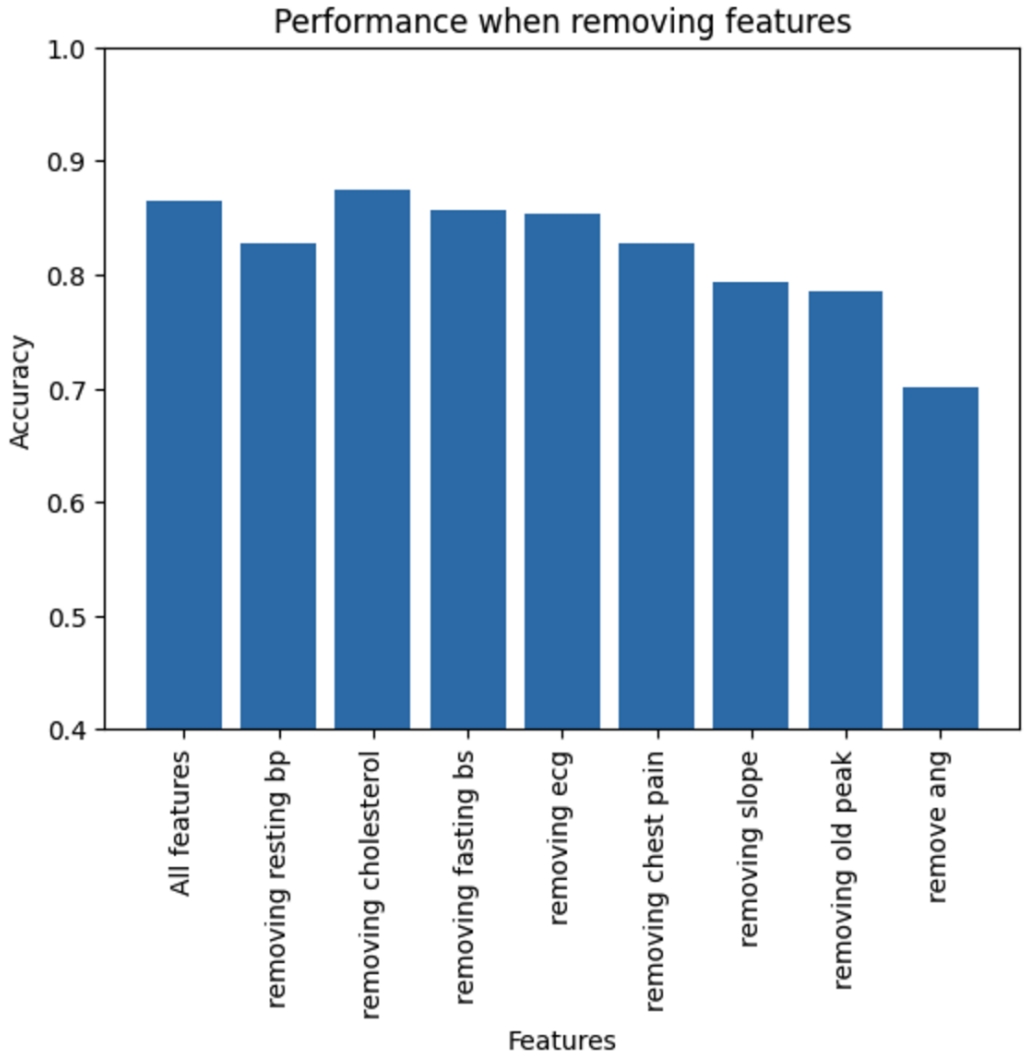


Figure 1: Feature reduction

4 Conclusion

In summary, feature importance showed signs that some features were not contributing very much information to the model. However, when running feature reduction these impacts were not as I expected. I would have assumed removing the less important features would have little impact, but removing the main features would have more of an impact than the charts show. Even though my chosen model showed signs of overfitting it yielded some of the highest valid accuracy along with lowest loss. However, given more time, a more appropriate choice given the smaller sample size and lower amount of features, would probably be a single layer model with some hyperparameter tuning.