Lab session: Regression and Classification Answers

Experiment 1: Heart rate estimation using regression techniques

P1. If you want to select a set of features manually, which features would you choose and why?

Features should be selected according to:

- High correlation of every feature with the target
- Low correlations among features
- The importance of the features according to our prior knowledge about the subject (for instance, we know that steps frequency is an important biomechanical feature for speed estimation)

P2. Do the automatically selected features match your manually selected features? Explain the reasons for any similarities and/or differences.

Reasons for similarities:

- Considering correlation of features among themselves as well as with the target

Reasons for differences:

- Considering multi-dimensional correlation among features and the target (which can be done by LASSO)
- Considering the biomechanical importance of the features according to our prior knowledge

P3. Do you think that the feature selection was useful in this exercise? Why?

Generally, feature selection is useful. However, we usually need to test several feature selection methods and fuse their results with our knowledge of the filed (e.g., biomechanical importance of the features) to obtain the best possible set of features.

P4. How do you interpret mean, std, and rmse errors of the models?

- Mean of error: it is the bias (accuracy) of the model and shows on how much on average the model is accurate. The sign of the mean error shows if the model is over/under estimating on average.
- Std of error: This metric shows the precision of the model. In other words, it shows how accurate the model is for every individual in our population.
- RMSE: this metric somehow combines both accuracy and precision and provide a general error parameter.

To make it clearer, consider the following extreme situations:

- A model with low mean error and high std error: such model is accurate on average (for instance for studying a population), but it can happen that for a specific individual, it faces high error
- Model with high mean error and low std error: Such model is not accurate on average, but this situation is almost the same for all individuals in the population.

P5. Considering all conditions, which model will you finally choose to estimate heart rate? Why?

Some criteria to choose the final model are:

- Check over-fitting or under-fitting by comparing the test and training error

  - Over fitting: training error << test error
  - Under fitting: test error << training error

- Check performance metrics of the model:

- o Accuracy (mean error)
- o Precision (std)
- o Overall error (rmse)

- Computational complexity of the model (the lesser the better)

Experiment 2: Atrial fibrillation classification using conventional classification techniques

P6. If you want to select a set of features manually, which features would you choose and why?

Please refer to the answer of P1.

P7. Do the automatically selected features match your manually selected features? Explain the reasons for any similarities and/or differences.

Please refer to the answer of P2.

P8. Do you see any signs of overfitting and/or underfitting of the models? Why?

Please refer to the answer of P5.

P9. Considering all conditions, which model will you finally choose to detect atrial fibrillation? Why?

Please refer to the answer of P5.