OBSS Assignment #2: Analysis of electromyogram of the uterus (EHG)

Matjaž Mav (63130148)

January 10, 2021

We used the sample entropy to analysed the electromyogram records of the uterus. Later we also calculate median frequencies and used all of the available features to learn the SVM classifier where we obtained validation accuracy of 0.888. Additionally we applied random forests to estimate feature importances.

1 Introduction

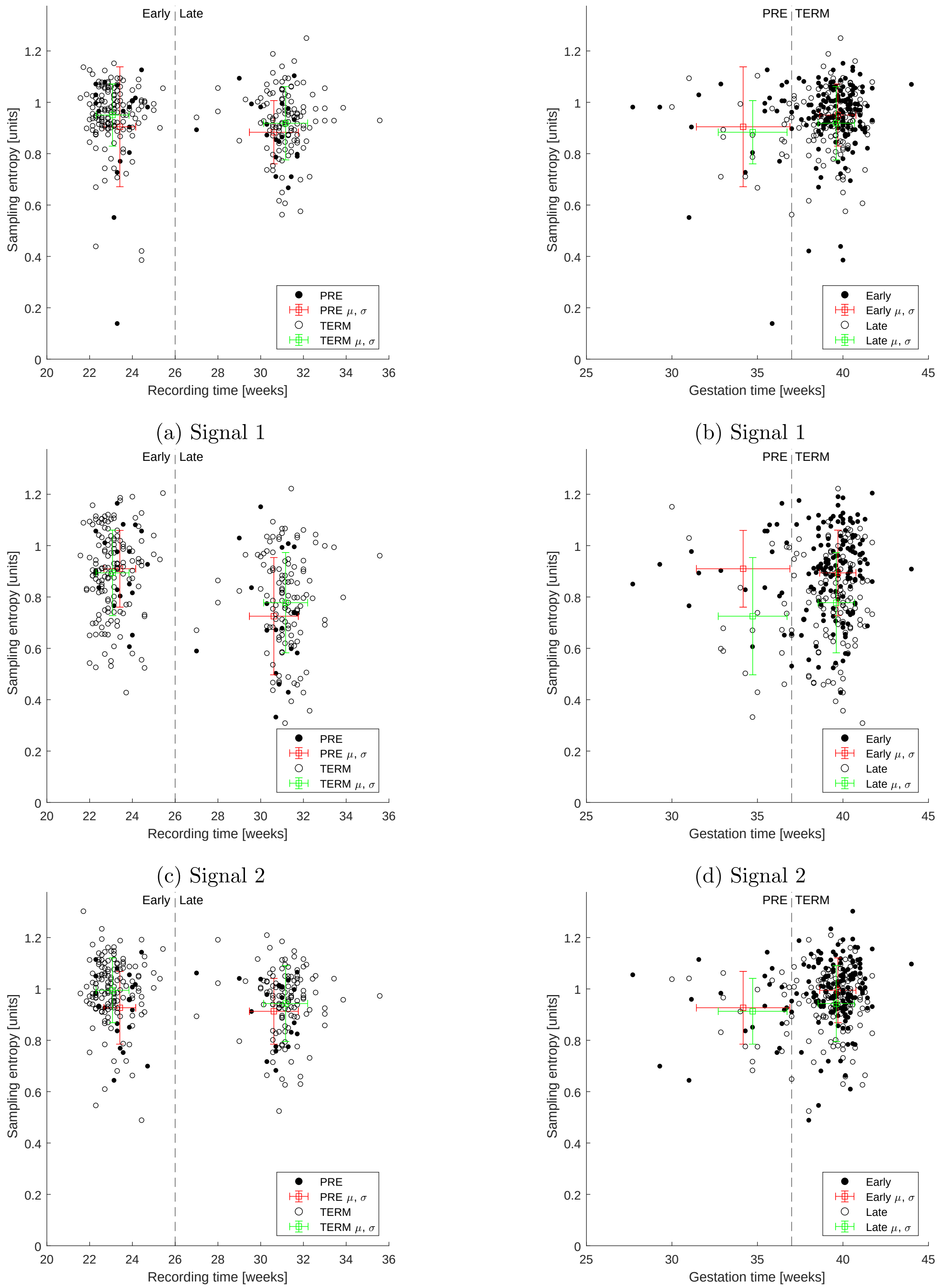
In this assignment we got familiar with sample entropy [1] and how to apply it to the EHG record of the uterus in order to predict preterm (PRE; $<$ 37 week) or term (TERM; $>=$ 37 week) delivery. Sample entropy is a measurement of regularity and estimation of extent to which the data did not arise from random process. So less predictable time series gives higher sample entropy. Analysis was performed on the TPEHGDB [2] dataset from the PhysioNet [3].

2 Methods

We performed analysis on all three signals that were already preprocessed with the Forward-Backward Butterworth filter of 4th order with the band-pass of 0.3 - 4 Hz. In our analysis we ignored the first 180 seconds of the signal since these interval contain transient effects of the preprocessing filters.

For all the record and signals in the dataset we computed sample entropy (pattern length $m = 3$, standard deviation $r = 0.15$) and the median frequency. Then we visualized (Figure 1) sample entropy in respect to the recording time (left column) and also to the gestation time (right column).

We observed that the sample entropy is higher for the early records (right column) and the term records (right column). This could mean that closer to the labour contractions get more regular and signal more predictable.



(a) Signal 1 (b) Signal 1

(c) Signal 2 (d) Signal 2

(e) Signal 3 (f) Signal 3

Figure 1: Sample entropy compared to the recording time (left column) and to the gestation time (right column) for all three signals. Raw signal was preprocessed with the Forward-Backward Butterworth filter of 4th order with the band-pass of 0.3 - 4 Hz.

3 Results

We extracted all metadata along with the median frequency and sample entropy and used in to estimate accuracy of the SVM classifier for the single signal. First we removed rows containing missing values, then mapped categorical data to numerical and then removed mean and scaled to unit variance each of the attributes.

We have evaluated two models (1) SVM-W where we applied proportional higher class weights to the preterm delivery and (2) SVM without weighting with the leave one out cross validation. Results are listed in the Table 1.

Model	μ_{acc}	σ_{acc}
SVM-W	0.692	0.462
SVM	0.888	0.316

Table 1: Leave one out cross validation accuracy

We also used random forest to estimate feature importances. It is clearly visable from the Figure 2 that the sample entropy is the most important feature in our dataset.

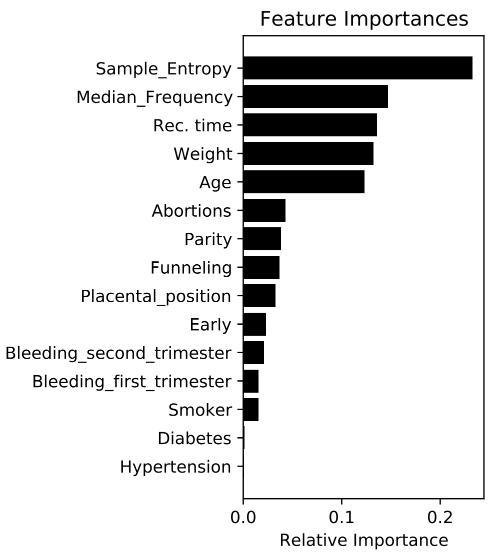


Figure 2: Feature importances obtained from the random forest classifier.

4 Conclusion

We are surprised that with class weighting we obtained lower validation accuracy. Also, it would be interesting to see how the SVM classifier would perform on not only a single signal but on all three signals that are available for each record.

References

[1] D. E. Lake, J. S. Richman, M. P. Griffin, J. R. Moorman, Sample entropy analysis of neonatal heart rate variability, American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 283 (3) (2002) R789–R797.

[2] G. Fele-Žorž, G. Kavšek, Ž. Novak-Antolič, F. Jager, A comparison of various linear and non-linear signal processing techniques to separate uterine emg records of term and pre-term delivery groups, Medical & biological engineering & computing 46 (9) (2008) 911–922.

[3] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals, circulation 101 (23) (2000) e215–e220.