

2nd assignment – Mortality prediction

Introduction to Clinical Data Science, 2022b

1 Preface

- In this exercise you will create a prediction model for mortality in the ICU using the MIMIC data set.
- Assuming you were successfully granted permission to the MIMIC-IV data. And assuming you have a GCP account (if not, please create one for free).
- Follow the instructions here:
<https://mimic.mit.edu/docs/gettingstarted/cloud/> in order to access the MIMIC-IV data from GCP BigQuery. More instructions can be found here:
<https://mimic.mit.edu/docs/gettingstarted/cloud/bigquery/>
- Make sure it works by running a simple query in BigQuery console.
- Follow steps on Colab notebook:
[https://colab.research.google.com/drive/11M3P7fWLsOjPgYOM9eFKCwmbibYRVke ?usp=sharing](https://colab.research.google.com/drive/11M3P7fWLsOjPgYOM9eFKCwmbibYRVke?usp=sharing) for prediction mortality using diagnosis data.
- Create a new Colab notebook in which you will follow similar steps, but this time, you will predict mortality during admission using laboratory data and demographic features.

2 Models' Learning

Create the next models:

1. Predict mortality during admission by number of labs
2. Predict mortality during admission by number of labs, gender, age, admission type, admission location, insurance, marital status, and ethnicity.
3. Predict mortality during admission by number of labs, gender, age, admission type, admission location, insurance, marital status, ethnicity and one hot encoding for m most common labs. Where $m \in \{8,16,32,64,128,256,512\}$.
4. Choose the optimal m and add to that model the most common 256 diagnoses (transformed to one hot encoding).
5. **Bonus:** (10 points): Same features as previous one, but limit the data that was collected in the first 24 hours of admission.

For each model, plot and report precision-recall curve and AUC, and ROC curve and AUC on the evaluation set.

3 Part 2 - Discussion

Add a discussion section. Discuss the differences between the models. Do you think the ICU can use your model and take advantage of it? Explain your answer and describe why it might be useful or useless.

4 Notes

- Your submission should include a URL for the Colab notebook. Make sure you run it all, and all plots and results are printed, so the exercise can be graded without re-running the analysis.
- Use the forum on Moodle for discussion, questions and answers.