SI 670 Term Project Proposal
11/10/2022
Elishua Shumpert and Steve Toner

Motivation
We propose to use regression and classification machine learning techniques to predict stay duration and mortality rates, respectively, among ICU patients with pneumonia.

Methods
Our two-fold approach will use both regression and soft classification techniques as appropriate:

- Stay Duration (Regression): Regularized linear regression, Artificial Neural Networks, Random Forest Regression, Kernel Ridge Regression, and Boosting
- Patient Mortality (Soft Classification): Logistic regression, soft-margin SVM, Gaussian Mixture Models

Datasets
Our data comes from the eICU Collaborative Research Database. The dataset consists of 200,000 admission records from major hospitals between 2014 and 2015; it includes features such as vital signals, admission diagnosis, treatment administered, admission and discharge times, and patient discharge status.

Evaluation
We will assess the performance of the methods described previously in reference to baseline dummy learners and “simple” techniques such as regularized linear and logistic regression. Success will be measured by improvement of algorithm performance beyond that of the baseline learning techniques.

Computing
We have submitted requests to Great Lakes for access to computing resources; we will utilize these resources for running larger / more complex models and deep learning methods.

Existing Work
While access to the full dataset is restricted to credentialed researchers, the eICU dataset is incredibly rich in detail and has been the subject of several previous research papers that apply machine learning to predict patient outcomes. However, we have seen most prominent papers focusing almost exclusively on a narrow class of deep learning models[1], namely LSTM-Graph Neural Networks (GNN). We will expand on prior work by exploring alternative learning methods and focusing specifically on patients admitted with pneumonia, in addition to applying traditional ML approaches.

Responsibilities of Each Group Member

- Elishua: Data pre-processing, EDA, linear regression, KNN
- Steve: Random Forest, Boosting, soft-margin SVM, GMM, ANN, error analysis

We plan to work together on feature engineering and writing the final report.

References

[1] Rocheteau, E., Tong, C., Veličković, P., Lane, N., & Liò, P. (2021, January 11). *Predicting patient outcomes with graph representation learning*. arXiv.org. Retrieved November 10, 2022, from https://arxiv.org/abs/2101.03940

[2] Danilatou V;Nikolakakis S;Antonakaki D;Tzagkarakis C;Mavroidis D;Kostoulas T;Ioannidis S; (n.d.). *Outcome prediction in critically-ill patients with venous thromboembolism and/or cancer using machine learning algorithms: External validation and comparison with scoring systems*. International journal of molecular sciences. Retrieved November 10, 2022, from https://pubmed.ncbi.nlm.nih.gov/35806137/