**Logistic Regression for Septic Shock Prediction**

*Introduction to Computational Medicine: The Physiome*

**Vivek Booshan, Jay Luo, Parimala Vedula, Elizabeth Zuerblis**

**Introduction**

Early prediction of potentially fatal septic shock is of enormous clinical value (**Angus et al., 2001; Liu et al., 2019**). This project investigates the classification of septic shock using static and dynamic linear models to estimate the probability that a patient develops septic shock using multiple clinical features. All code used for this project is based on the instructor-provided MATLAB datasets and scripts.

**Constructing the Generalized Linear Model (GLM)**

**Data Exploration**

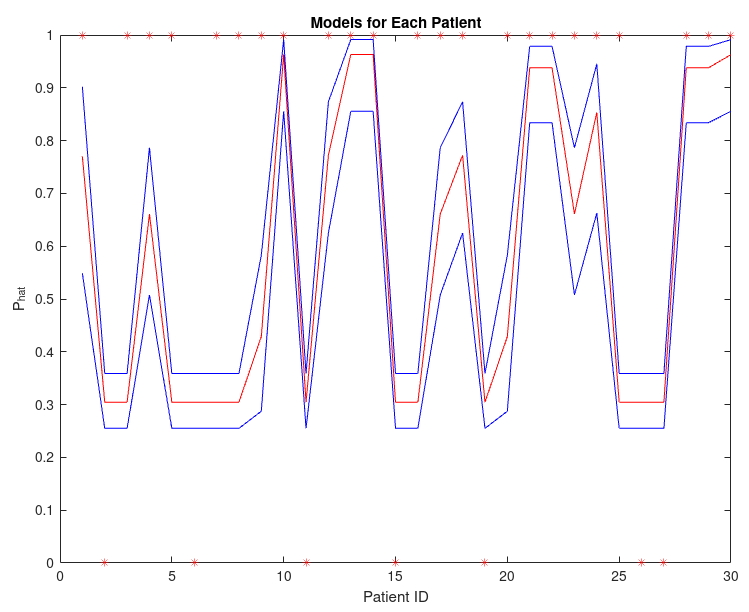
A GLM was used to determine the probability that a given patient has sepsis or not. The log-odds of entering septic shock can be written as a linear combination of learned model parameters (alphas) and the dataset features (x’s), where the first term (alpha\_0) is the intercept term (1).

(1)

The following variables from the clinical dataset were used as features in the GLM model to separate sepsis from non-sepsis patients. The static features are as specified in the project assignment PDF.

**Static Model**

After training the GLM, our static model performs well with an AUROC of around 0.79. We chose to exclude gender and age in our model predictions as this resulted in a greater AUROC (i.e. more robust classifications) than including all static features.



***Figure 1: p\_hat (predicted probabilities) with 95% confidence intervals for first 30 patients of static training set***

**Learned parameters: static model**

Our trained parameters with 95% confidence are (LB = lower bound; UB = upper bound):

| **Parameter** | **Coefficient estimate (LB, Beta, UB)** |
| --- | --- |
| Intercept term | (-1.0721, **-0.8265**, -0.5809) |
| Respiratory Comorbidities | (1.5832, **2.0480,**  2.5128) |
| Heart Comorbidities | (0.1643, **0.5399**, 0.9154) |
| Infection | (1.1030, **1.4940**, 1.8850) |

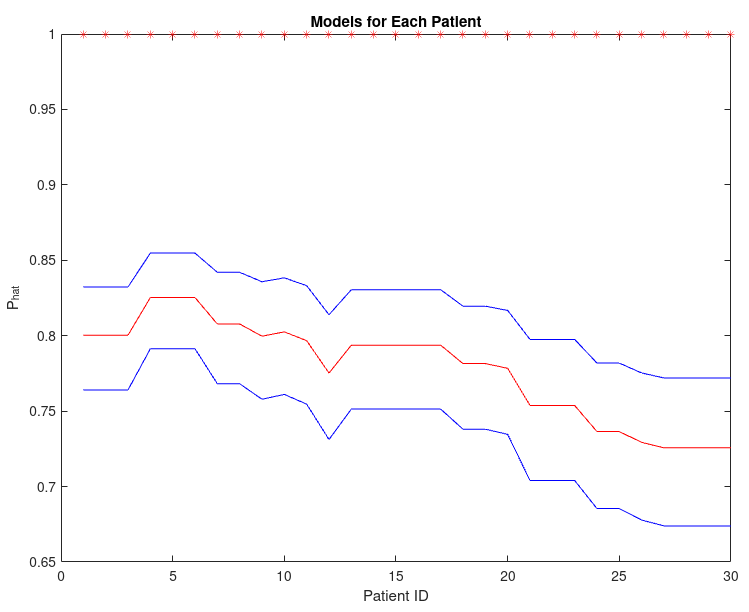
**Dynamic Model**

For the dynamic GLM, the following equation was used where p(t) is the probability the patient has sepsis at time t as a function of the patient features (x\_i) (2).

(2)

We chose not to exclude any dynamic variables, following Liu et al. (2019)’s conclusion of the importance of these dynamic features. We also believe all of these variables to be physiologically relevant and important for generalizable dynamic GLMs for septic shock prediction.

The decision rule for Phat uses Matlab’s OPTROCPT, which provides the optimal operating point of an ROC by finding the slope of the cost of misclassification.

****

***Figure 2: predicted septic shock probabilities (p\_hat) values (by dynamic model) for dynamic training set***

**Learned parameters: dynamic model**

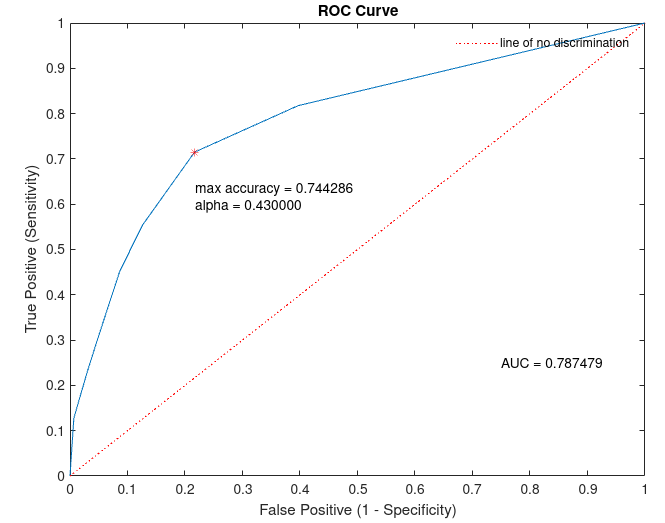
Our trained parameters with 95% confidence are (parameters as in project assignment PDF):

| **Parameter** | **Coefficient estimate (LB, Beta, UB)** |
| --- | --- |
| Intercept term | (-0.8276, **-0.7550**, -0.6824) |
| [**Static**] Respiratory Comorbidities | (1.8157, **1.8443**, 1.8728) |
| [**Static**] Heart Comorbidities | (0.0260**, 0.0506,** 0.0752) |
| [**Static**] Infection | (0.8641, **0.8890,** 0.9138) |
| [**Dynamic**] Lactate | (-0.1071, **-0.1016**, -0.0962) |
| [**Dynamic**] Cardiovascular SOFA score | (-0.5241, **0.5374,** 0.5507) |
| [**Dynamic**] GCS | (-0.0513, **-0.0475**, -0.0437) |
| [**Dynamic**] Heart Rate | (0.0176, **0.0182**, 0.0189) |
| [**Dynamic**] Partial pressure of oxygen | (-0.0021, **-0.0019**, -0.0017) |
| [**Dynamic**] Fraction of inspired oxygen | (0.0085**, 0.0091,** 0.0097) |

**Evaluating the GLMs**

**Static Model**

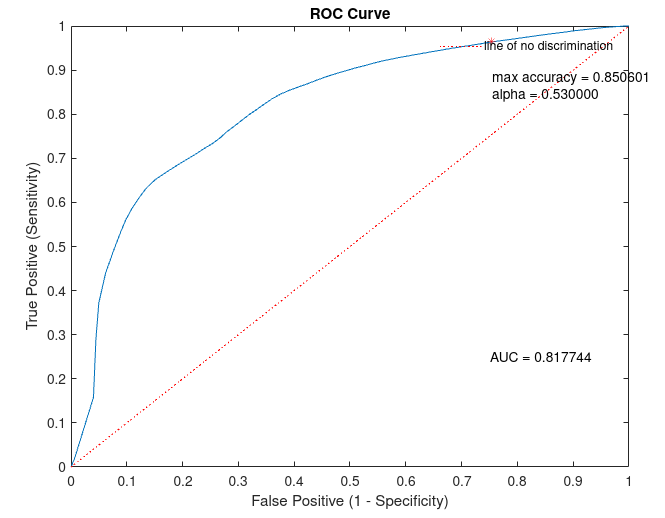
Our static GLM achieved an AUROC of ~0.79 for the static **training** set.

****

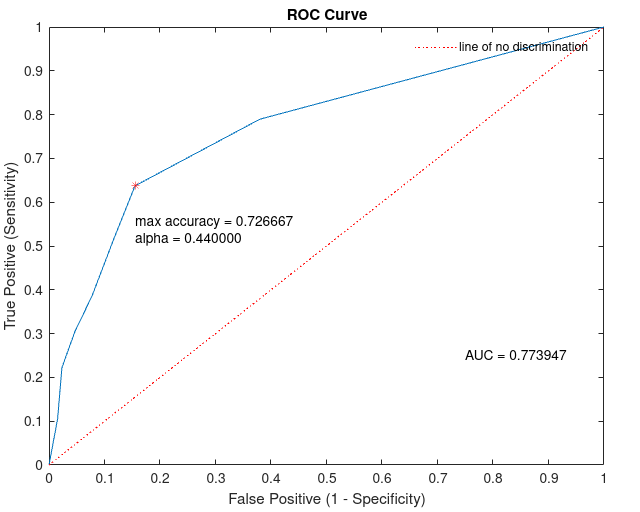
***Figure 3: ROC curve for the static training set***

**Dynamic Model**

Our dynamic GLM achieved an AUROC of ~0.82 for the dynamic **training** set.

****

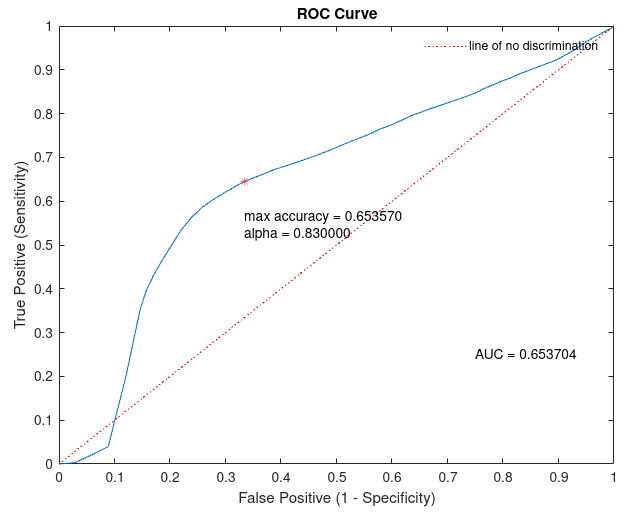
***Figure 4: ROC curve on the dynamic training set***

**Model Validation**

**Static Model**

We achieved an AUROC of around 0.77 on the static validation set (n = 300 patients). This value is just slightly below the AUROC of ~0.79 achieved with our static training set (n = 700 patients), which is expected as the model is predicting on ‘unseen’ validation data. In fact, this shows our model, trained on a selected static feature set, generalizes decently to unseen data.

***Figure 5: Static model ROC curve for static validation data***



**Dynamic Model**

We achieved a respectable AUROC of ~0.65 on the dynamic validation set. This is worse than the value achieved on the dynamic training set (~0.82), which we expect since this validation set (as in the static case) was not seen by the model during training.

***Figure 6: Dynamic model ROC curve for dynamic validation data***

**Conclusions**

We have developed static and dynamic GLMs to model the probability of patients entering into septic shock. Our models demonstrated strong performance on both the training and validation sets (more so for the static model). Interestingly, we found the respiratory comorbidities and infection static variables to be influential features in our predictions for both the static and dynamic models, demonstrated by their high positive ‘betas’/weights. GLMs of septic shock are therefore useful statistical models that may provide clinically actionable, interpretable, and early predictions of septic shock.

**Contributions**

**Vivek Booshan:**, coding, plotting, writing

**Jay Luo:** writing, coding

**Parimala Vedula:**

**Elizabeth Zuerblis:** writing

**References**

Angus, D. C., Linde-Zwirble, W. T., Lidicker, J., Clermont, G., Carcillo, J., & Pinsky, M. R. (2001). Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Critical Care Medicine*, *29*(7), 1303–1310. <https://doi.org/10.1097/00003246-200107000-00002>

Liu, R., Greenstein, J. L., Granite, S. J., Fackler, J. C., Bembea, M. M., Sarma, S. V., & Winslow, R. L. (2019). Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. *Scientific Reports*, *9*(1), Article 1. <https://doi.org/10.1038/s41598-019-42637-5>