Predicting Patient Survival Time using Deep Survival Analysis

Sidhika Balachandar, Daniel Michael, Sumer Sao

CS221 Final Project

Stanford Computer Science

Background

Standard Survival Analysis

Problem Statement

- Given patient clinical data, can we predict when a patient will die?

Demographics | Medical History | Laboratory Data | Procedure Information | Medications

Event Prediction Model

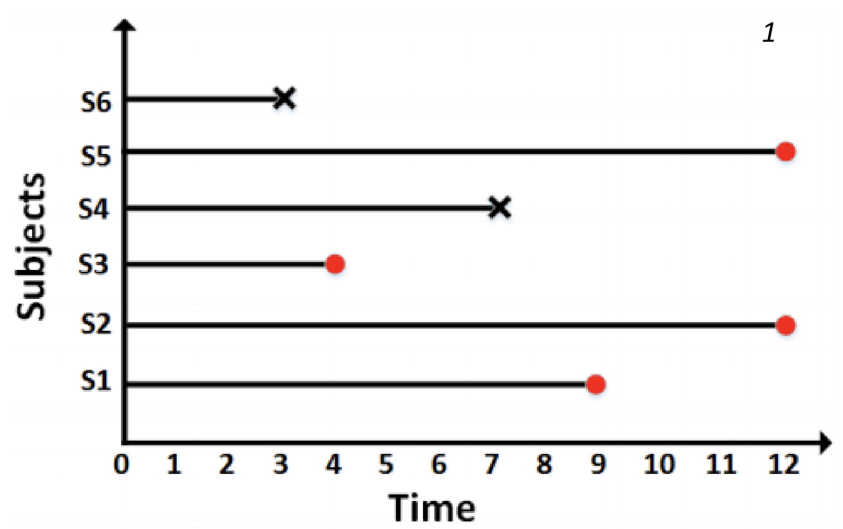
Impact:
Lower healthcare costs
Improve quality of life

Overview

- **Inputs**: clinical data on patients with Primary Biliary Cirrhosis
  - Features include drug, age, sex, stage of disease, etc.
- **Output**: Survival function - probability of death within t days
  - Enables prediction of time range of death with ϵ certainty

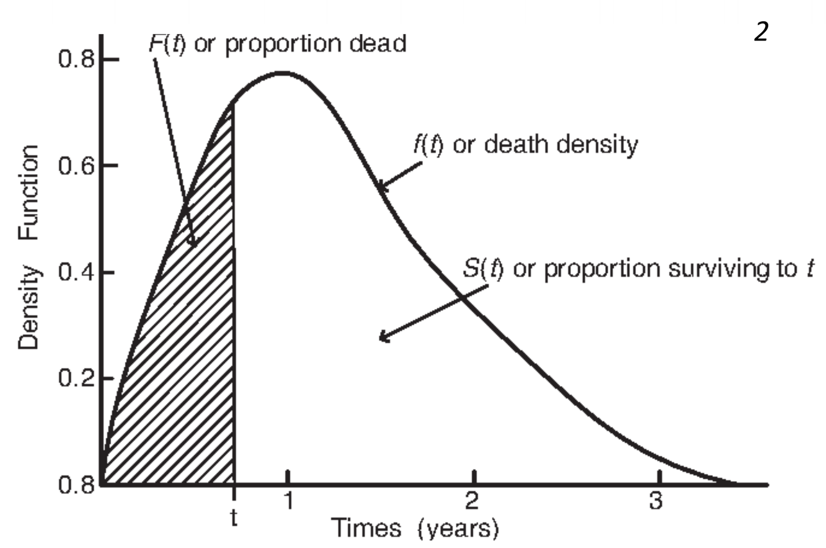
Right Censored Data:

- Why censor?
  - Patient lives for entire duration of study (S2, S5)
  - Patient drops out (S1, S3)
- 54% of our data was censored



Survival Analysis Basics:

- S(t) = survival function (our output)
- $h(t) = \frac{f(t)}{S(t)}$ = hazard function = probability of death at next instant



Evaluation Metrics:

- How can we evaluate our model if we output a survival function?
- Concordance-index (AUC) measures discriminative ability of a model
- Why C-index?
  - provides **relative** risk between patients
  - Helpful in comparing similar patients
- C-index = ratio of the true positive rate to the false positive rate

1 Chandan K. Reddy and Yan Li, "A Review of Clinical Prediction Models", in Healthcare Data Analytics, 2015.
2 Ping Wang, Yan Li, Chandan, K. Reddy, "Machine Learning for Survival Analysis: A Survey". ACM Computing Surveys, 2017.

Methods

Cox Proportional Hazards Model

- Why CPH?
  - CPH can accommodate for censored data
  - CPH outputs a survival function
- Process (used scikit-survival algorithms)
  - Estimates weights (β) for each feature and initial hazard (h_0)
  - β estimated with linear regression of log hazard on each patient's data
  - h_0 estimated with maximum likelihood estimator
  - Use β and h_0 to calculate the survival function $S(t)$

$$S_0(t) = exp(-h_0(t))$$

$$S(t|X_i) = S_0(t) \times exp(X_i\beta)$$

  - Regularizations used:
    - LASSO (L1 Norm)
    - Ridge (L2 Norm)
    - Elastic Net (convex combination of LASSO and Ridge)

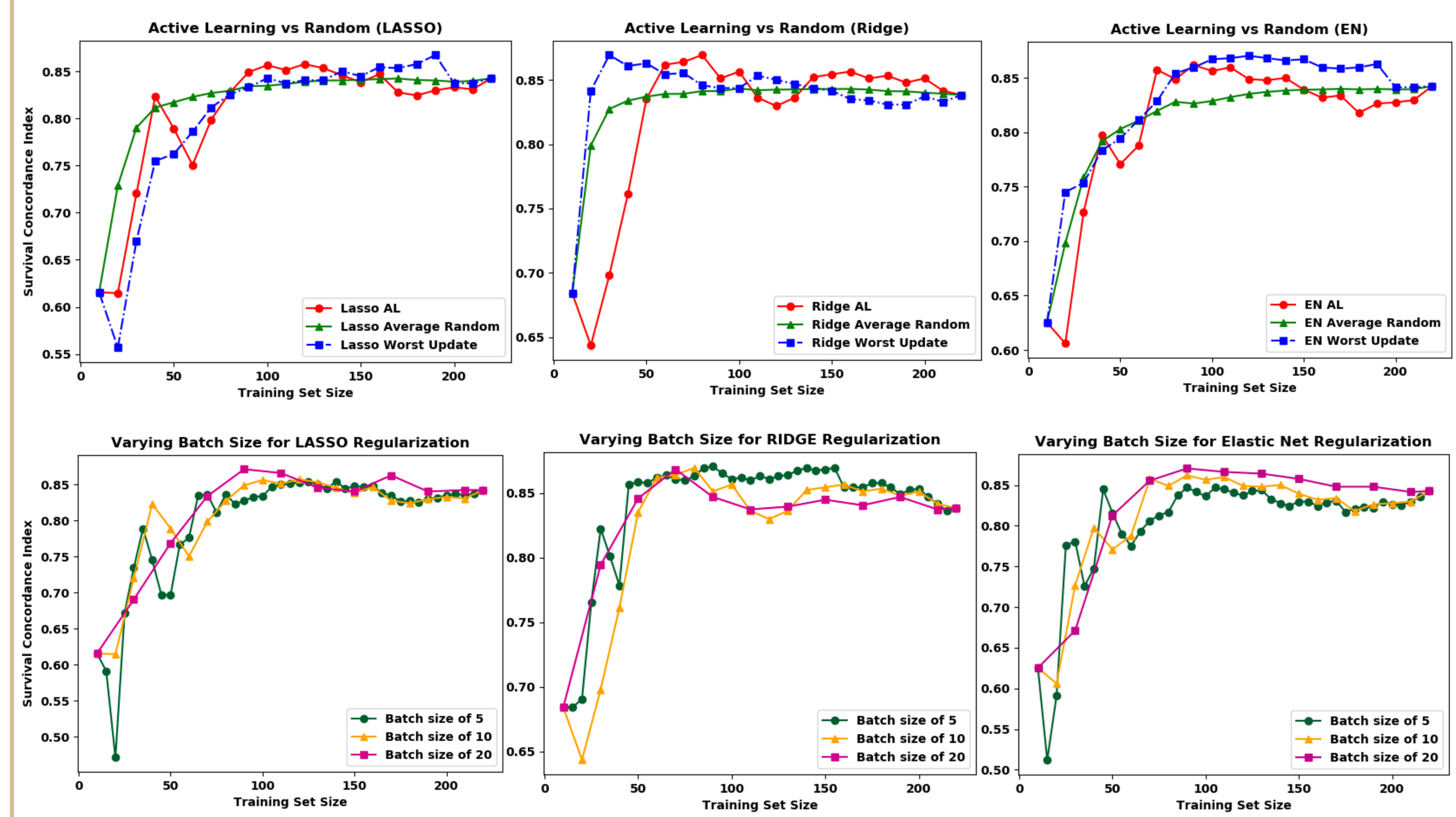
Active Learning

- Why AL?
  - Problem: Labeling a patient (dead/alive) can take years
  - AL identifies whether a patient would be helpful for the model
- What does AL do?
  - Identifies most representative data
  - Higher learning rate (similar accuracy with less training data)
- Process (hand coded the algorithm)
  - See which unlabeled data point contributes the most
  - Ask oracle for label of this point to be included in the data set.
  - Re-run our CPH model with this new data set.

Results

Model	C-Index
Cox (LASSO)	0.843
Cox (Ridge)	0.838
Cox (Elastic Net)	0.842
AL (LASSO)	0.843
AL (Ridge)	0.838
AL (Elastic Net)	0.842

Results



Conclusions

- CPH without AL can predict the survival function with high accuracy
- Out of all CPH only models LASSO regularization does the best
- Active learning has the exact same accuracy as the CPH model
- In most cases active learning has a higher learning rate
- All batch sizes have similar accuracies
- Our PCB dataset is not the best dataset for testing Active Learning
  - Because increase in learning rate is marginal and worst-case updates are comparable to best case
  - We used this dataset because it was the most accessible

Future Work

- Run on more clinical dataset
  - For example from TCGA
- Integrate clinical and genomic data for survival analysis
  - Having gene expression data, for genes like tumor suppressing TP53, can increase accuracy of model
  - Incorporate Active Learning framework to integrated model
- Work with an MD at Stanford Hospital
  - More recent and relevant data
  - See our work implemented in practice