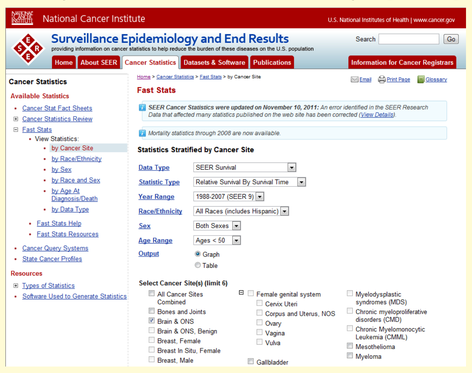
Learning Patient-Specific Cancer Survival Distributions as a Sequence of Dependent Regressors

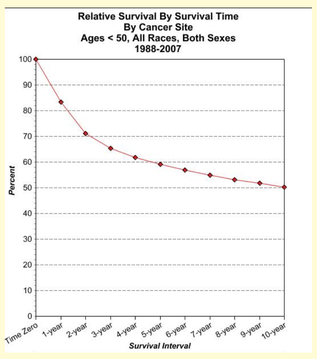
Chun-Nam Yu[1] Russ Greiner[1] Hsiu-Chin Lin[1] Vickie Baracos[2]
Dept of Computing Science[1]/Dept of Oncology[2], University of Alberta, Canada

UNIVERSITY OF ALBERTA
ALBERTA INNOVATES CENTER FOR MACHINE LEARNING

Introduction

- Cancer patients and doctors need to make decisions about treatments and end-of-life care
- Good prognostic models help manage the uncertainty and make better decisions
- Use statistics such as *median survival time*, *5 year survival rate*
- Now: estimated from a large heterogenous population, based on only:
  - Site + Stage of cancer
  - (Perhaps) 1 or 2 covariates (e.g., age, sex, race)





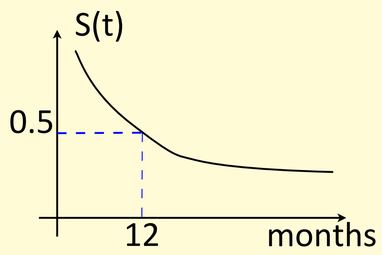
- Should use other information already in electronic health records:
  - prescriptions
  - blood test results
  - performance assessment by physicians
- e.g. the Alberta Cancer Registry
- **Our GOAL:** Build accurate patient-specific prognostic predictors with electronic health records

Basic Survival Analysis

- Survival function:

$$S(t) = P(T \geq t)$$

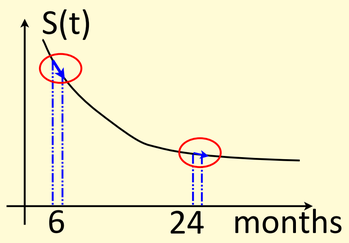
is the proportion of patients surviving longer than t months



- Hazard function:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t}$$

is the instantaneous rate of failure/death at time t, given the patient survives longer than t



- Censored Observations: Some patients leave in the middle of a study or remain alive at the end of a study ⇒ only a lower bound on their survival time.

Survival Regression Models

- One of the most important regression model in survival analysis is the Cox Proportional Hazards (PH) Model [1]
- Given covariates $\vec{x}$, Cox regression assumes the hazard function have the form

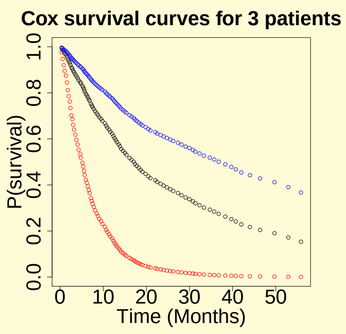
$$h(t \mid \vec{x}) = h_0(t) \exp(\vec{\theta}^T \vec{x}),$$

where $h_0(t)$ is a base hazard independent of $\vec{x}$.

- The hazard ratio is constant (independent of t) for two patients $\vec{x}_1$ and $\vec{x}_2$:

$$h(t \mid \vec{x}_1)/h(t \mid \vec{x}_2) = \exp(\vec{\theta}^T(\vec{x}_1 - \vec{x}_2))$$

- Objective depends on rank of survival time only; cannot handle covariates with time-varying effects
- Other survival regression models include Aalen linear hazards model [2], parametric survival regression based on Weibull, log-normal distributions



- **Survival Analysis VS Survival Prediction:**

	Survival Analysis	Survival Prediction
Goal:	Evaluate prognostic factors and treatment effectiveness	Predict accurate patient-specific survival time
Focus:	Evaluation on populations	Evaluation on individuals
Uses:	Testing new drugs & medical devices	Treatment planning & patient management

Data - Cancer Patient Composition and Patient Attributes

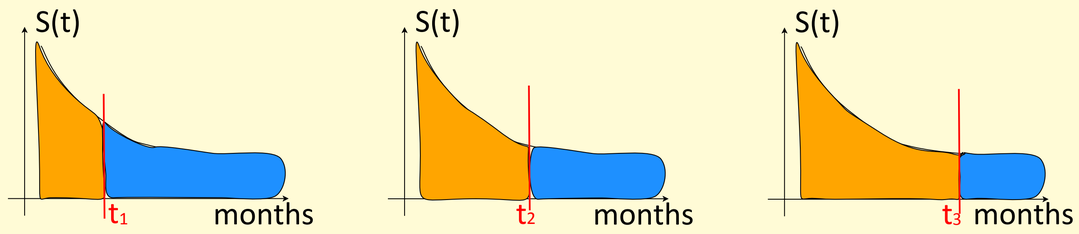
site\stage	1	2	3	4
Bronchus & Lung	61	44	186	390
Colorectal	15	157	233	545
Head and Neck	6	8	14	206
Esophagus	0	1	1	63
Pancreas	1	3	0	134
Stomach	0	0	1	128
Other Digestive	0	1	0	77
Misc	1	0	3	123

basic	age, sex, weight gain/loss, BMI, cancer site, cancer stage
general wellbeing	no appetite, nausea, sore mouth, taste funny, constipation, pain, dental problem, dry mouth, vomit, diarrhea, performance status
blood test	granulocytes, LDH-serum, HGB, lyphocytes platelet, WBC count, calcium-serum, creatinine, albumin

Predicting Personalized Survival Distributions

- Consider a simpler problem: will a subject with covariate $\vec{x}$ survive $\geq t$ months?
- Model this with a simple logistic regression model:

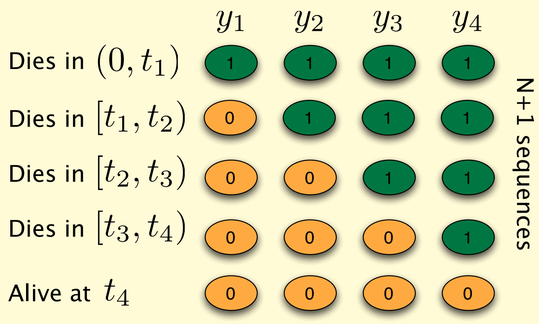
$$P_{\vec{\theta}}(T \geq t \mid \vec{x}) = 1/(1 + \exp(\vec{\theta}^T \vec{x} + b))$$

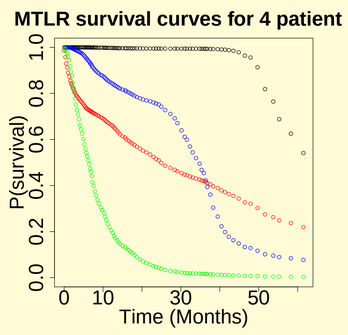


- Multiple thresholds $t_1, t_2, \ldots$ to capture survival function at different times

$$P_{\vec{\theta}_j}(T \geq t_j \mid \vec{x}) = 1/(1 + \exp(\vec{\theta}_j^T \vec{x} + b_j))$$

- Predict a sequence of dependent bits





- **Multi-task logistic regression** for survival prediction:
  - Likelihoood of observing a sequence:

$$P(Y = (y_1, y_2, \ldots, y_N) \mid \vec{x}) = \frac{\exp(\sum_{j=1}^{N} y_j(\vec{\theta}_j \cdot \vec{x} + b_j))}{Z(\vec{\Theta}; \vec{x})}$$

  - Add multi-task regularizer to prevent overfitting:

$$\min_{\vec{\Theta}} \frac{C_1}{2} \sum_{j=1}^{N} \|\vec{\theta}_j\|^2 + \frac{C_2}{2} \sum_{j=1}^{N-1} \|\vec{\theta}_{j+1} - \vec{\theta}_j\|^2 - \sum_{i=1}^{n} \sum_{j=1}^{N} \left[y_j(s_i)(\vec{\theta}_j \cdot \vec{x}_i + b_j) - \log Z(\vec{\Theta}; \vec{x}_i) \right]$$

- Similar to CRF training objective, but
  - no transition features
  - no sharing of node potentials
  - Partition function $Z(\vec{\Theta}; \vec{x})$ involves only a linear number of terms
- Model also related to local regression approaches [3]
- Advantage of the MTLR approach:
  - No PH assumption - effects of covariates can change with time
  - Handles censoring: by integrating out hidden variables in a survival sequence
  - More accurate survival probability predictions (below)

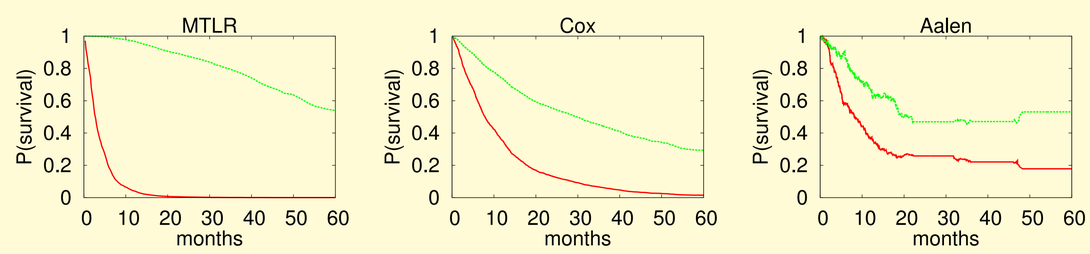
Experimental Results

- Classification Accuracy and Survival Rate Prediction (5CV):

Accuracy	5 month	12 month	22 month
MTLR	86.5(0.7)	76.1(0.9)	74.5(1.3)
Cox	74.5(0.9)	59.3(1.1)	62.8(3.5)
Aalen	73.3(1.2)	61.0(1.7)	59.6(3.6)
Baseline	69.2(0.3)	56.2(2.0)	57.0(1.4)

MSE	5 month	12 month	22 month
MTLR	0.101(0.005)	0.158(0.004)	0.170(0.007)
Cox	0.196(0.009)	0.270(0.008)	0.232(0.016)
Aalen	0.198(0.004)	0.278(0.008)	0.288(0.020)
Baseline	0.227(0.012)	0.299(0.011)	0.243(0.012)

- *Baseline* is the common approach of predicting with cancer site & stage only.
- Predicted survival curves for two test patients (Patient 1: 3 months, Patient 2: censored at 46 months)



- Optimizing clinically relevant loss functions (prediction p, true survival t):
  - Absolute error on survival time (AE): $|p - t|$
  - Absolute error on log survival time (AE-log): $|\log p - \log t|$
  - Relative absolute error (RAE): $\min\{|p - t/p|, 1\}$

	MTLR	Cox	Aalen	CSVR	Site+Stage
AE	9.58(0.11)	10.76(0.12)	19.06(2.04)	9.96(0.32)	11.73(0.62)
AE-log	0.56(0.02)	0.61(0.02)	0.76(0.06)	0.56(0.02)	0.70(0.05)
RAE	0.40(0.01)	0.44(0.02)	0.44(0.02)	0.44 (0.03)	0.53(0.02)

- *CSVR* is censored support vector regression [4].
- Similar results on two other large survival datasets, *SUPPORT2* and *RHC*

References

[1] *D.R. Cox, Regression models and life-tables, J ROY STAT SOC B MET, 1972*
[2] *O.O. Aalen, A linear regression model for the analysis of life times, STAT MED, 1989*
[3] *T. Hastie & R. Tibshirani, Varying-coefficient models, J ROY STAT SOC B MET, 1993*
[4] *Shivaswamy et al., A support vector approach to censored targets, ICDM 2007*