Latent Survival with Gaussian Processes

Alan Saul^{a,b}, Neil D. Lawrence^{a,b}

Survival Analysis

Survival analysis is an area of statistics concerned with the analysis of time-to-event data. It's most widespread application is in the analysis of clinical data. Here we propose a model for learning a latent space representation of patients using a survival analysis based likelihood for survival times and a Gaussian Process mapping to covariate measurements.

Hazard functions

A key component in many survival analysis techniques is the concept of a hazard function. Here rather than characterizing the non-negative event time, T, through its failure density function (p.d.f), we consider the instantaneous rate at which the event is occurring, known as the hazard rate,

$$h_T(t|\boldsymbol{\theta}) = \lim_{\delta t \to 0} \frac{\Pr(t < T \leq t + \delta t | T > t)}{\delta t}.$$

Hazard models often assume that the hazard itself can be broken down into two components, a baseline hazard that is independent of measurements of the subject, and a relative hazard component that causes a subjects hazard to deviate from the baseline, often multiplicatively. Such models are known as Proportional Hazards models.

Cox Proportional Hazards Model

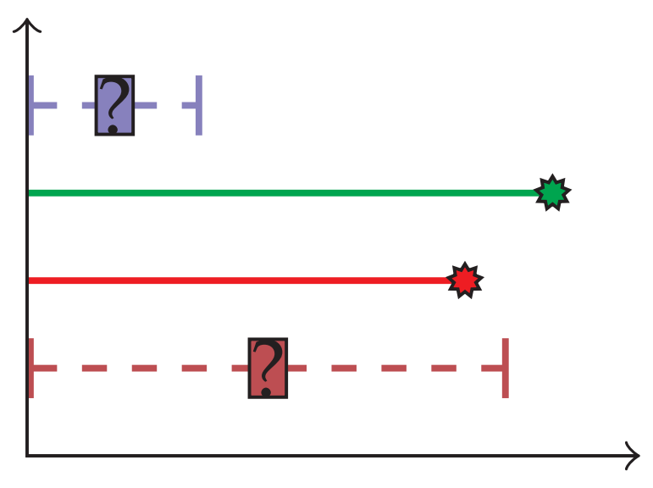
Perhaps the most widely used method in survival analysis is the Cox Proportional Hazards Model [1],

$$h_T(t|\boldsymbol{\theta}) = h_{0T}(t|\boldsymbol{\theta}) \exp(\boldsymbol{\beta}^\top \mathbf{x}_i).$$

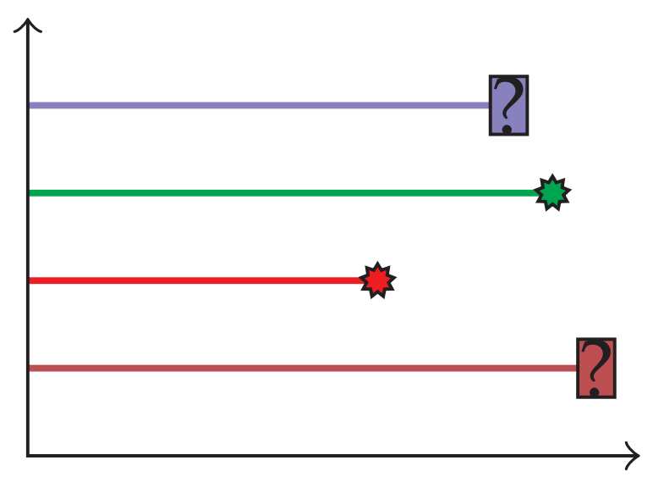
The baseline hazard which is left undefined in the model is multiplicatively increased or decreased depending on measurements of covariates of the subject, by a simple exponentiated linear component. Recent work [4, 2] has been focusing on relaxing the log-linear relationship that covariate measurements have in the proportional hazards model.

Censoring

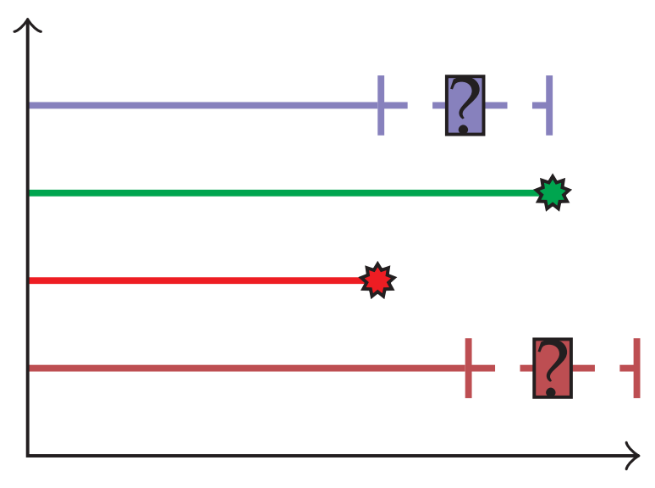
A key complication of the analysis of time-to-event data, particularly in a clinical setting is the presence of "censored" data. Censored data involves data that is partially known and must be included in the analysis so as not to introduce a bias.



Left censoring



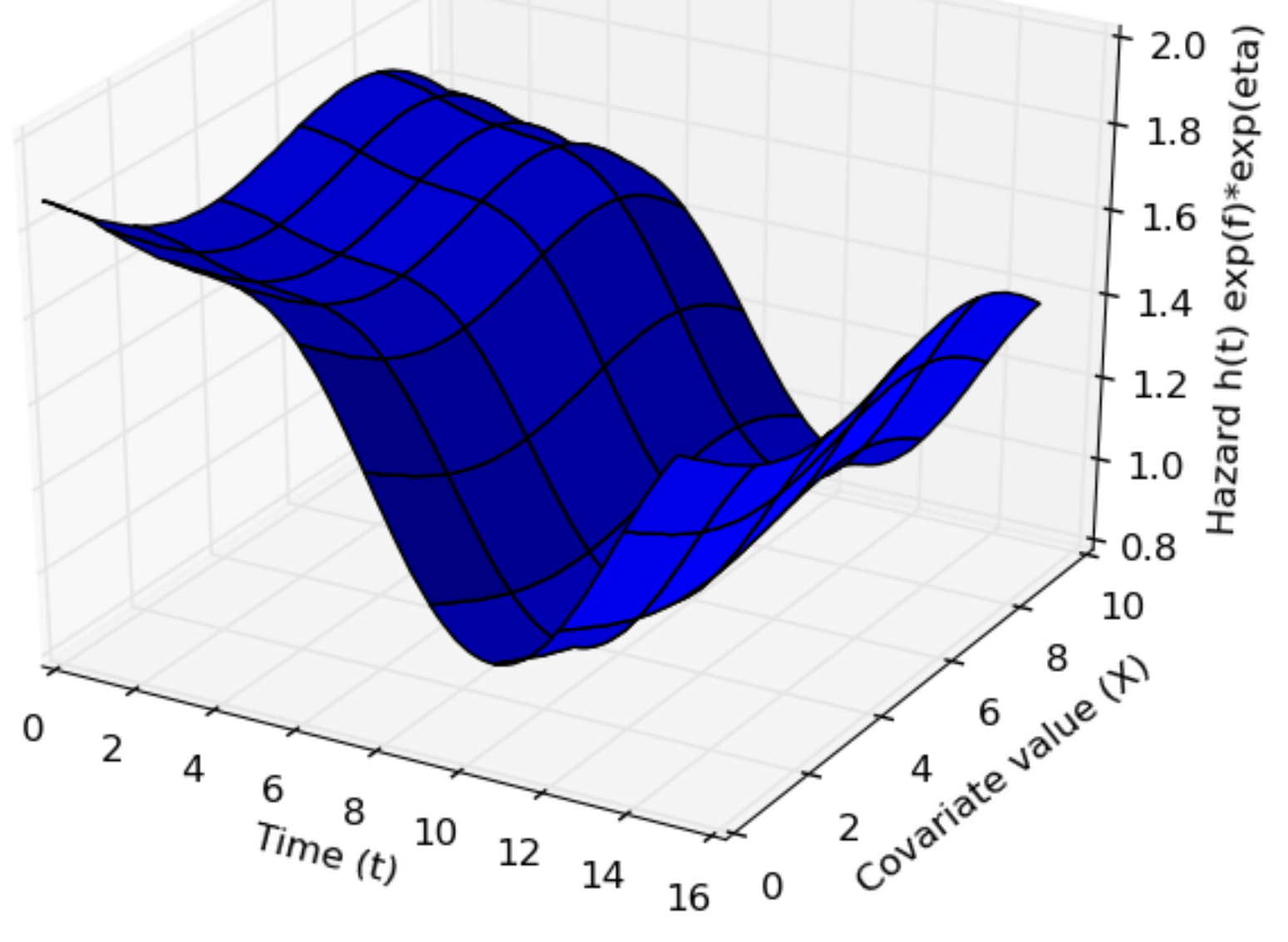
Right censoring



Interval censoring

Gaussian Process Survival Analysis

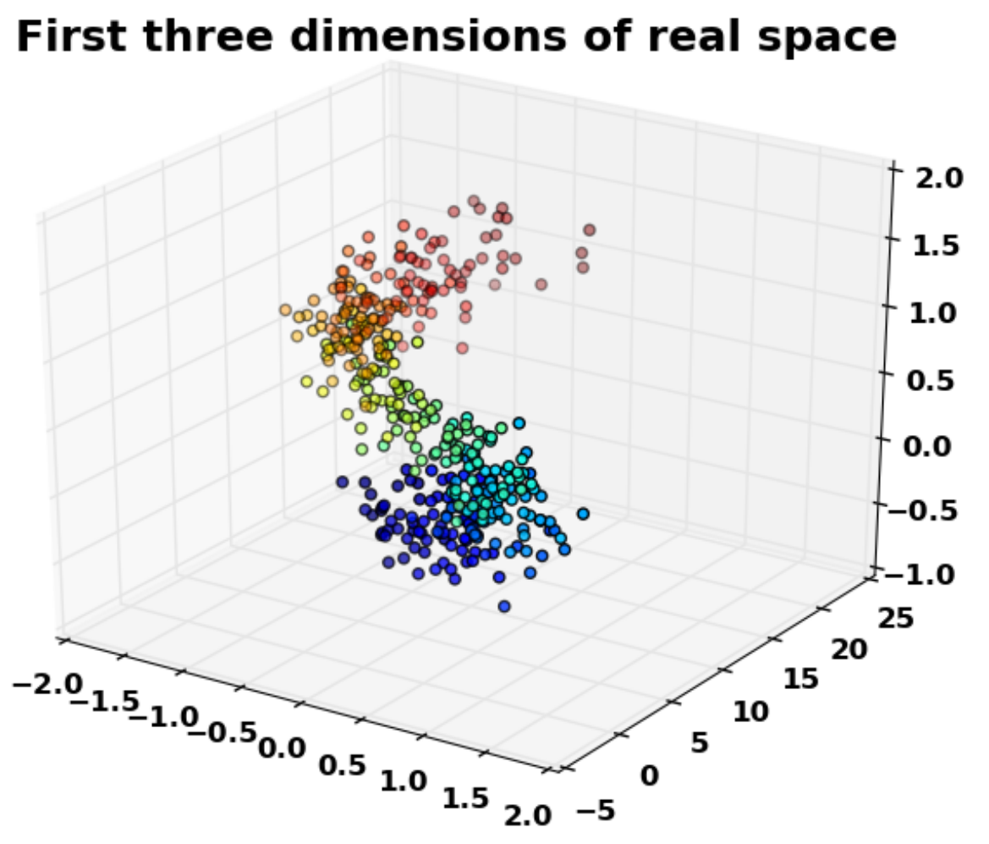
Joensuu et al. [2] produced a model that relaxes this log-linear assumption by reformulating the linear component as a Gaussian Process. This allows for enormous flexibility of the hazard function. The baseline hazard is also modelled as a Gaussian Process, providing the ability to make longitudinal predictions that the cox model would be incapable of making without further approximations. The likelihood assumes that the hazard is piecewise constant within chosen regions.

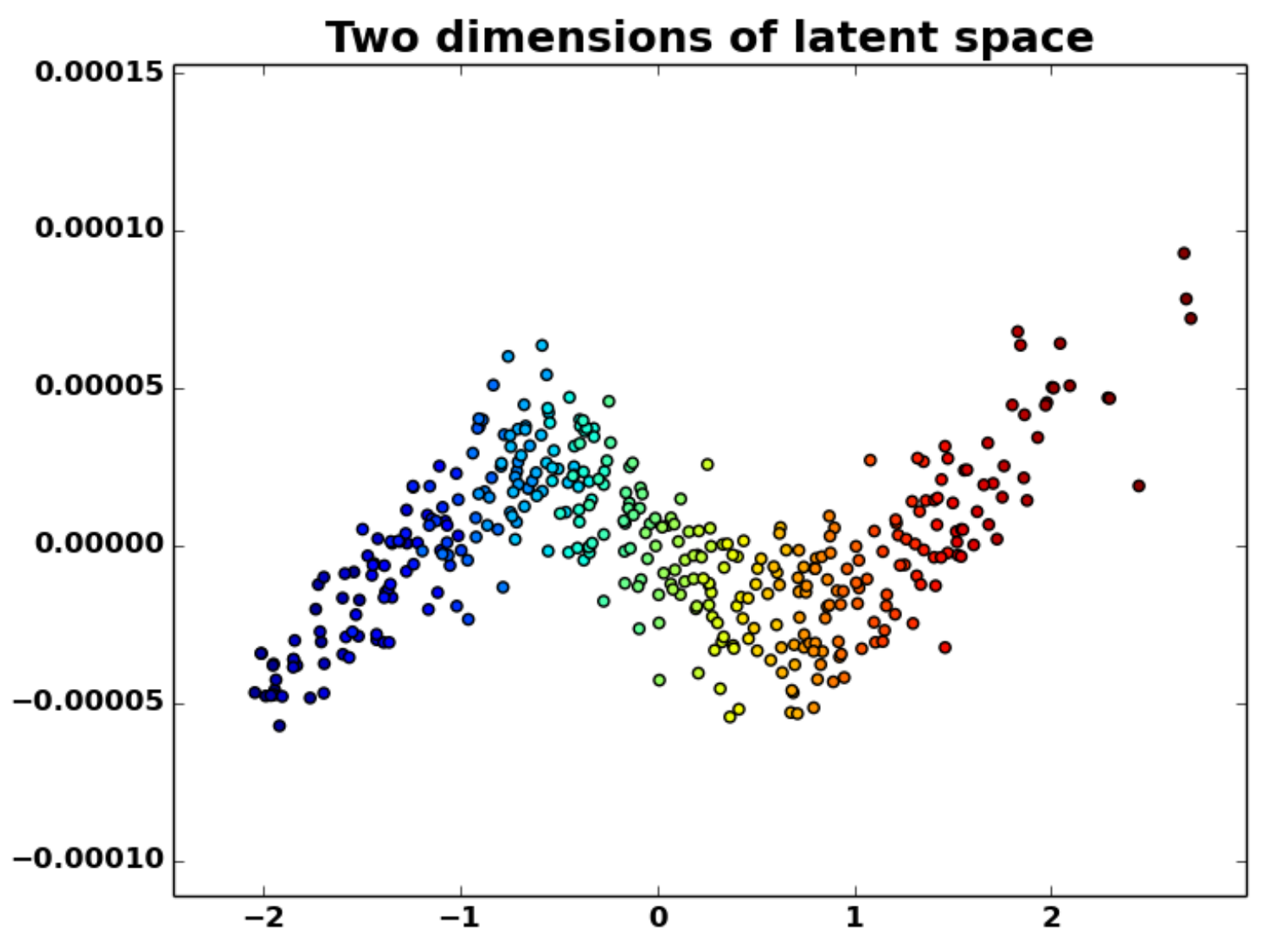


Example of flexible hazard function

Gaussian Process Latent Variable Model

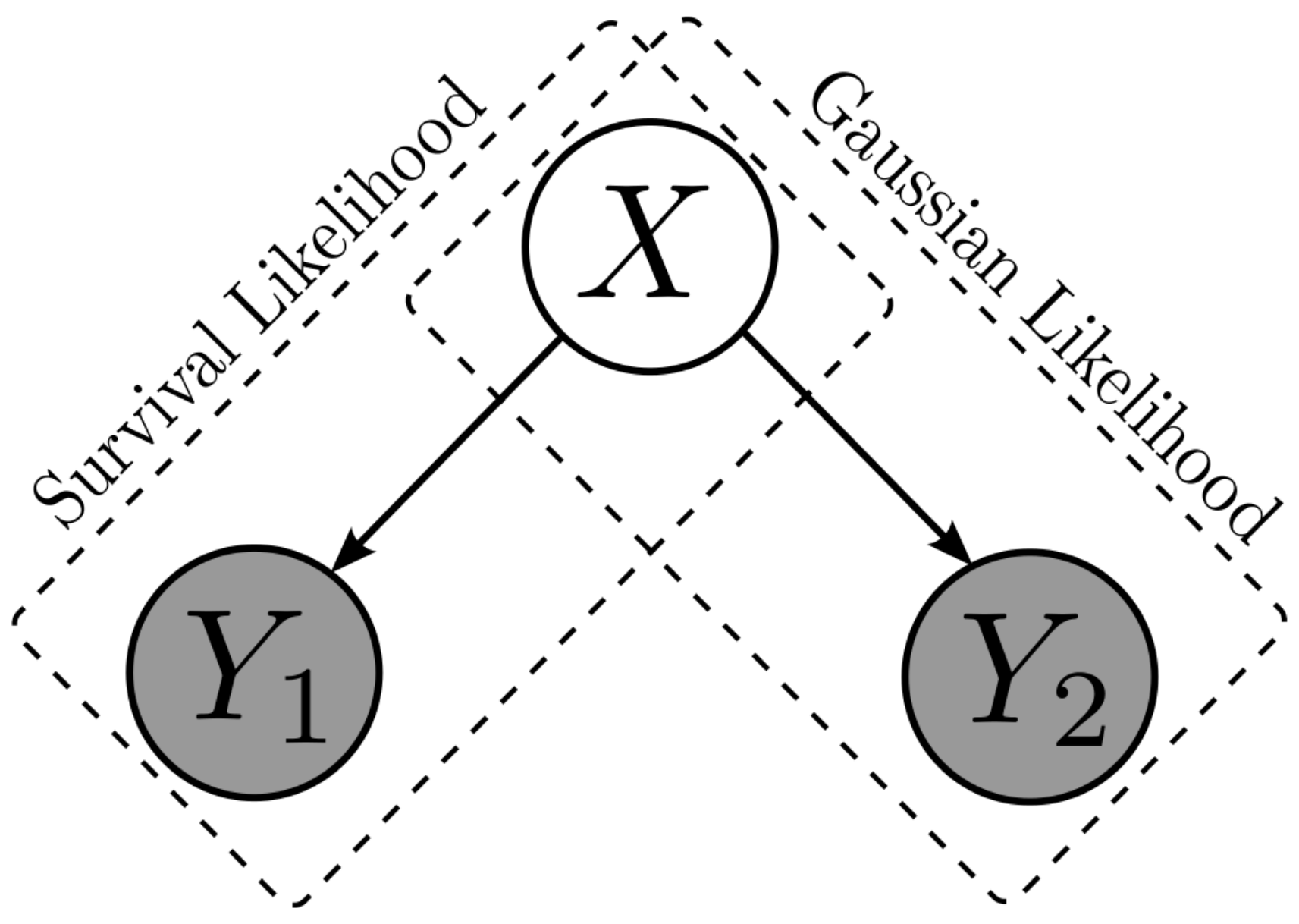
Probabilistic dimensionality reduction can be performed using a model known as the Gaussian Process Latent Variable Model [3]. The model allows for a non-linear mapping to be assumed from a lower dimensional input space to a higher dimensional output space, contrasting to the linear mapping Probabilistic Principle Component Analysis provides.





Proposed model

We propose a model that discovers an underlying latent space from complex data that lives on a higher dimensional manifold, such that survival times for 'similar' patients share similar survival times.

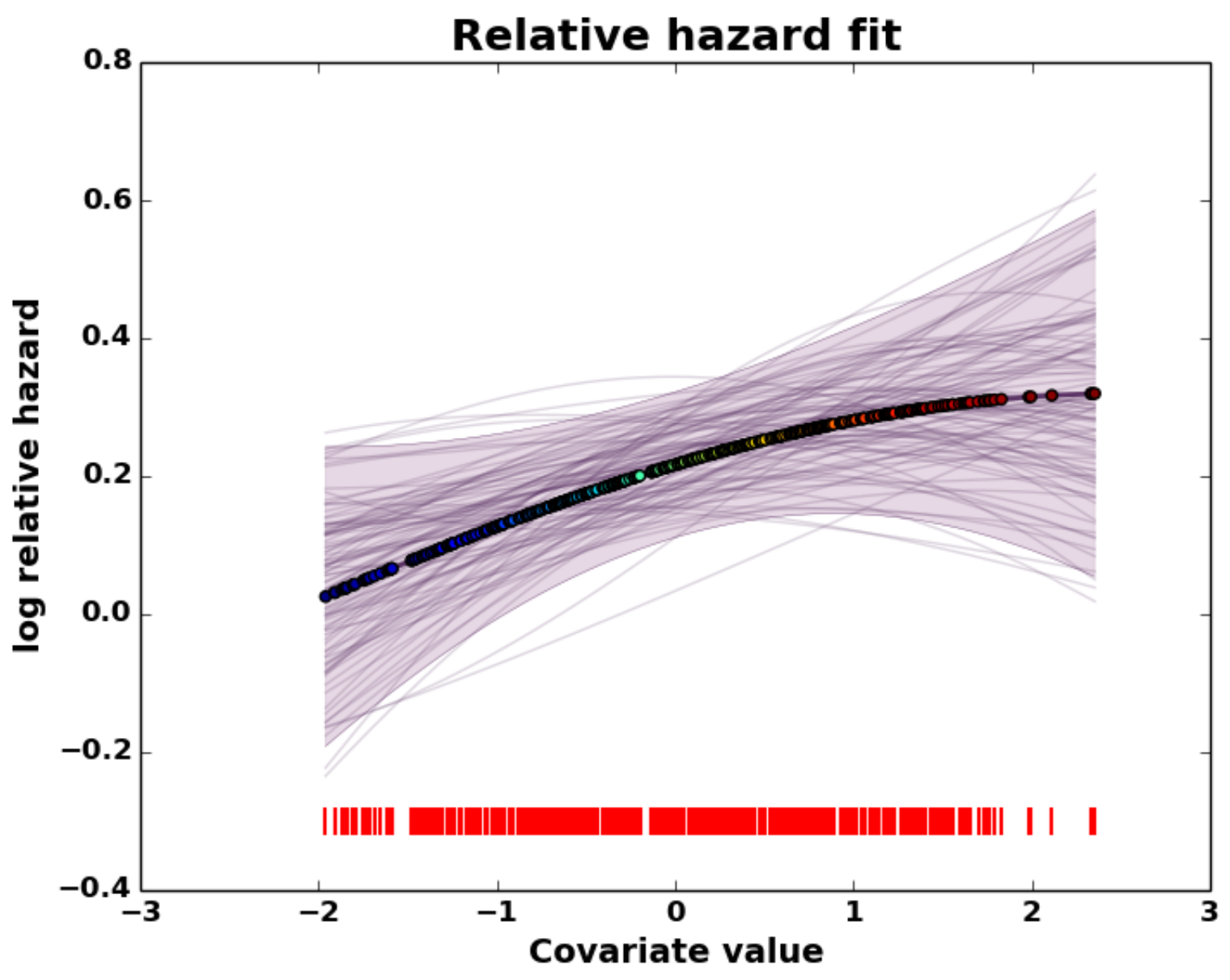


Joint GPLVM

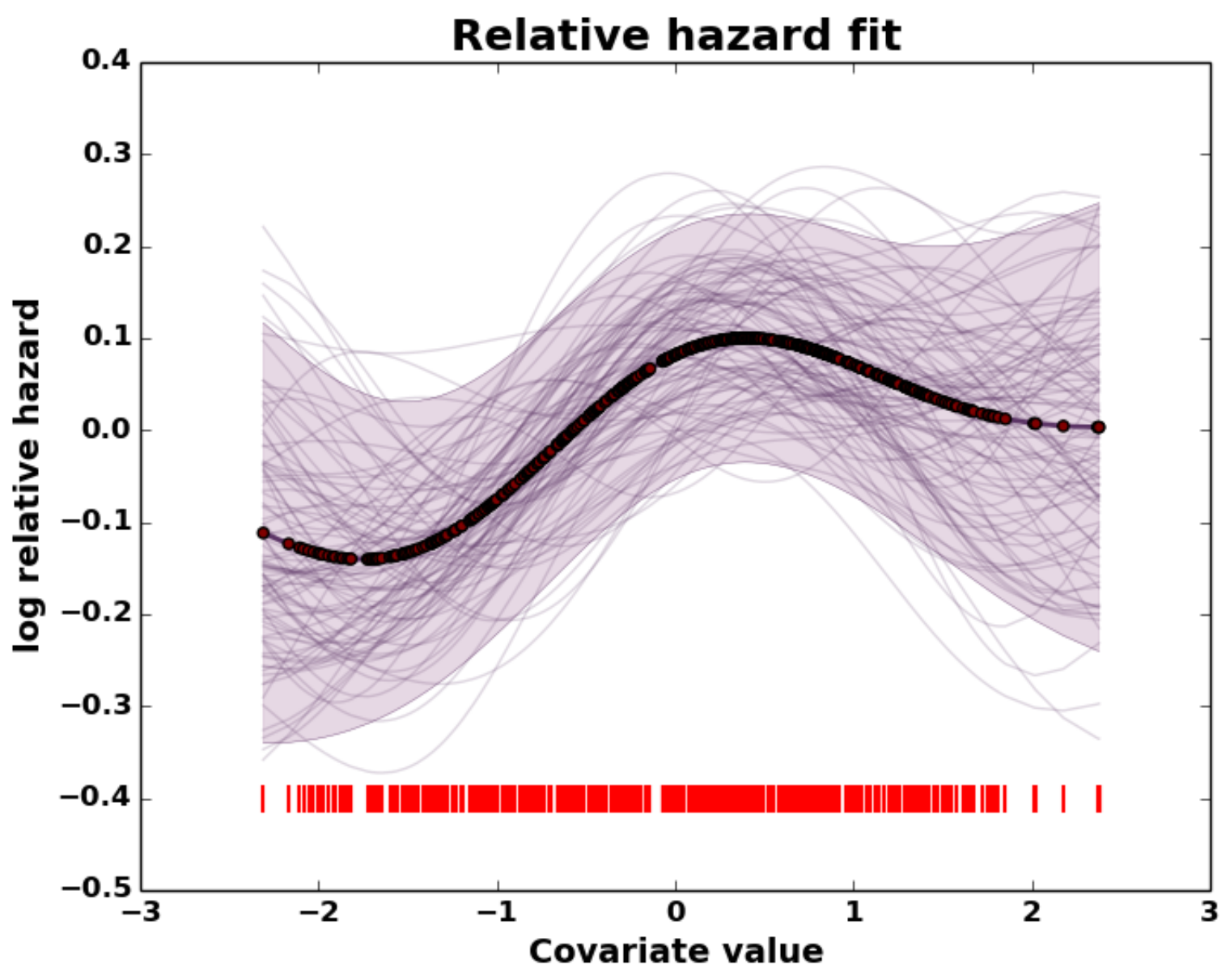
This provides a more general picture of similarity between patients, and potentially increase the predictive power during survival regression. Inclusion of survival data in the model will additionally guide the dimensionality reduction. We may be able to detect clusters of patients within the latent space, pointing towards confounding factors such as hospital status or genetic similarities.

Preliminary results

We show the potential advantages of allowing the survival times to affect the latent space by comparing the results of a model on a toy data-set, where the relative hazard mapping is known to be linear for the correct latent representation of the input data. It is the task to learn the latent space correctly such that this linear relative hazard is discovered.



Jointly learnt model



Independently learnt models

References

[1] D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society*, 34(2):187–220, 1972.

[2] H. Joensuu, A. Vehtari, J. Riihimäki, T. Nishida, S. E. Steigen, P. Brabec, L. Plank, B. Nilsson, C. Cirilli, C. Braconi, A. Bordoni, M. K. Magnusson, Z. Linke, J. Sufliarsky, M. Federico, J. G. Jonasson, A. P. Dei Tos, and P. Rutkowski. Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts. *The lancet oncology*, 13(3):265–74, Mar. 2012.

[3] N. Lawrence. Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models. *Journal of Machine Learning Research*, 6:1783–1816, 2005.

[4] S. Martino, R. Akerkar, and H. Rue. Approximate Bayesian Inference for Survival Models. *Scandinavian Journal of Statistics*, 38(3):514–528, 2010.

[a] University of Sheffield, Department of Computer Science, UK [b] Sheffield Institute for Translational Neuroscience (SITraN), UK