# High-dimensional, massive sample-size Cox PH regression for survival analysis

1. Introduction

* Previously: survival analyses limited to applications with only a handful of predictors and a few hundreds or thousands of observations
* Recent advances in data acquisition techniques & access to high computation power -> larger data for analysis
* Large-scale applications: medical adverse event monitoring, longitudinal clinical trials, business data mining tasks
* Require methods for analyzing high dimensional, massive sample-size (HDMSS) data in a survival analysis framework
* Cox model & regularized Cox models
* Works well for small-scale problems >< do not scale well to HDMSS data due to use of costly Newton-Raphson iterations that require inverting large matrices
* Possible workaround and approx. -> large estimated coef variances, numerical ill-conditioning and poor predictive accuracy / calibration
* To solve the optimization problem -> regularized Cox survival modeling that scales for HDMSS data
* Exploit a variation of the cyclic coordinate descent optimization technique
* Avoids overfitting, provide improved predictive performance, efficient during fitting and prediction time

1. Related work (see paper for more info)
2. Regularized Cox survival analysis

* Assuming typical survival analysis setting
* n = number of individuals in the training data
* Survival times where
* : indicator variable such that
* : p-vector of covariates for individual i
* Assume that are conditionally independent given and censoring is non-informative
* Observe data:
* : p-vector of unknown, underlying model parameters
* Assume that survival times arise in an independent and identically distributed fashion from density and survival functions and parameterized by
* Likelihood of the parametric model:

(3.1)

* Cox proportional hazard model – semi-parametric hazard function of the form:

(3.2)

* Similarly, survival function unfolds as:

(3.3)

* Through (3.2) & (3.3), (3.1) falls out as:

(3.4)

* In the absence of explicit specification of the baseline hazard -> hard to work with directly.
* Alternatively, Cox proposes to maximize the partial likelihood function

(3.5)