Homework 3 | Survival Analysis

50 points

Karen Hovhannisyan

To do

Dataset

- ID: Subscriber ID
- region: region code
- tenure: lifetime
- age: subscriber's age
- marital: subscriber's marital status
- address: number of years living in the same address
- income: Subscriber's annual income (K)
- ed: education level
- retire: retired (Yes/No)
- gender:
- voice:
- internet:
- forward: call forwarding
- custcat: customer category
- churn:

Parametric Models

- Build AFT models with all the available distributions:
  - for Python: visit here
  - for R: you can find in the slides
- Compare the models
- Visualize all the curves: one plot for all
- Which model would you use as a decision maker (think about other factors apart from the above comparisons)
- Keep significant features
- Keep the final model

CLV

Calculate CLV **per customer** based on the **final model**; you can use the same logic provided in the slides. Explore CLV within different segments.

Report

Write a short report (1-2 paragraphs) about your findings. The goal is to understand the factors affecting the churn risk.

- interpret the coefficients
- try to find the most valuable segments: describe the definition of being valuable according to you
- assuming the data represents the population, how much would be your **annual** retention budget? (hint, you should look at CLV, Survival probabilities and detect the number of at-risk subscribers within a year)
- what else would you suggest for retention?

Submission Rules

- Github Repo with:
  - Code
  - Markdown/Notebook: report and code
  - requirements.txt file (in case of Python)
  - Readme.md file with brief intro
  - No manual uploads!