DEEP LEARNING FOR PATIENT-SPECIFIC KIDNEY GRAFT SURVIVAL ANALYSIS

Group Members:

Nishat Zaman (011 151 264)
Nawshiba Tasnim Ahmed (011 161 145)
Ferdous Zaman (011 162 125)
Ebnul Mahmood Shovan (011 162 058)

INTRODUCTIONS

In previous presentation we didn't understood deep learning, survival function, hazard function, cox proportional hazard model etc. in this presentation we will discuss details about those topics.

Dataset

Table 2: **Characteristics of the real-life datasets.**

Datasets	Nb. (%) cens.	Nb. unique t (%)	Nb. inst.	Nb. feat.	Missing val.
aids2	1082 (38.1)	1013 (35.6)	2843	12	no
colon death	477 (51.3)	780 (84.0)	929	37	yes
colon recurrence	461 (49.6)	749 (80.6)	929	37	yes
flchain	5705 (72.5)	2977 (37.8)	7874	23	yes
mgus 2 tgt2	421 (30.4)	272 (19.7)	1384	5	yes
nwtco	3457 (85.8)	2767 (68.7)	4028	9	no

Dataset

age = age in years, sex = F=female, M=male, sample.yr = the calendar year in which a blood sample was obtained, kappa = serum free light chain, kappa portion, lambda = serum free light chain, lambda portion, flc.grp = the FLC group for the subject, as used in the original analysis, creatinine = serum creatinine, mgus = 1 if the subject had been diagnosed with monoclonal gammapothy (MGUS), futime = days from enrollment until death. Note that there are 3 subjects whose sample was obtained on their death date, death = 0=alive at last contact date, 1=dead, chapter = for those who died, a grouping of their primary cause of death by chapter headings of the International Code of Diseases ICD-9.

Dataset

Shape (7874, 12)

	Unnamed: 0	age	sex	sample.yr	kappa	lambda	flc.grp	creatinine	mgus	futime	death	chapter
0	1	97	F	1997	5.70	4.860	10	1.7	0	85	1	Circulatory
1	2	92	F	2000	0.87	0.683	1	0.9	0	1281	1	Neoplasms
2	3	94	F	1997	4.36	3.850	10	1.4	0	69	1	Circulatory
3	4	92	F	1996	2.42	2.220	9	1.0	0	115	1	Circulatory
4	5	93	F	1996	1.32	1.690	6	1.1	0	1039	1	Circulatory

```
data.describe()
data.chapter.describe()
```

```
count            2169
unique             16
top       Circulatory
freq              745
Name: chapter, dtype: object
```

Dataset

	age	sex	kappa	lambda	flc.grp	creatinine	mgus	futime
0	97	0	5.70	4.860	10	1.7	0	85
1	92	0	0.87	0.683	1	0.9	0	1281
2	94	0	4.36	3.850	10	1.4	0	69
3	92	0	2.42	2.220	9	1.0	0	115
4	93	0	1.32	1.690	6	1.1	0	1039

Dataset

```
kmf.event_table
```

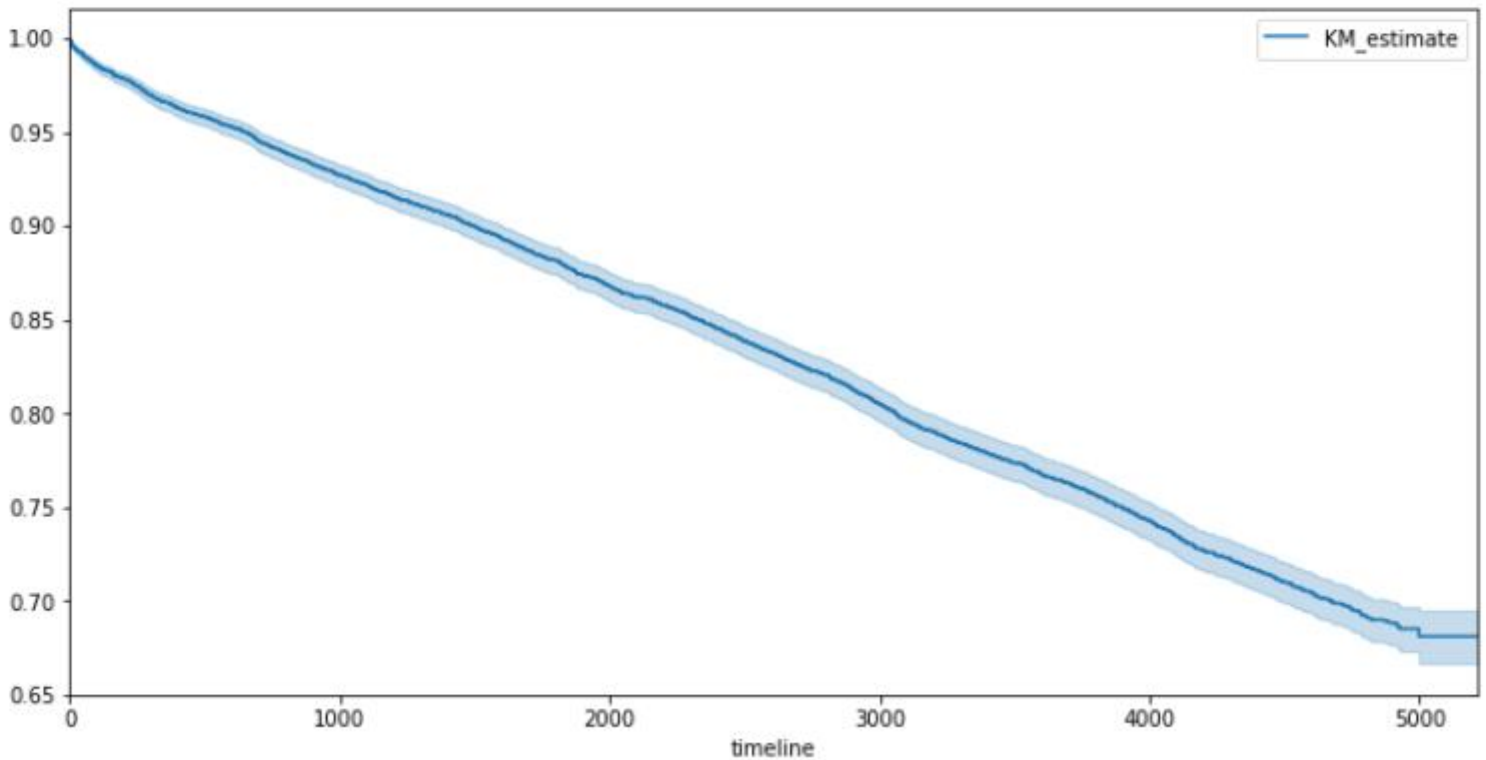
event_at	removed	observed	censored	entrance	at_risk
0	3	3	0	7874	7874
1	8	4	4	0	7871
2	1	1	0	0	7863
3	6	4	2	0	7862
4	3	3	0	0	7856
5	3	2	1	0	7853
6	5	3	2	0	7850
7	5	3	2	0	7845
8	6	4	2	0	7840
9	3	2	1	0	7834

Dataset

The *removed* column contains the number of observations removed during that time period, whether due to death (the value in the *observed* column) or censorship. So the *removed* column is just the sum of the *observed* and *censorship* columns. The *entrance* column tells us whether any new subjects entered the population at that time period. Since all the players we are studying start at $time = 0$ (the moment they were drafted), the *entrance* value is 15,592 at that time and 0 for all other times.

The *at_risk* column contains the number of subjects that are still alive during a given time. The value for *at_risk* at $time = 0$, is just equal to the *entrance* value. For the remaining time periods, the *at_risk* value is equal to the difference between the time previous period's *at_risk* value and *removed* value, plus the current period's *entrance* value. For example for $time = 1$, the number of subject's *at risk* is 10,995 which is equal to 15,592 (the previous *at_risk* value) - 4,597 (the previous *removed* value) + 0 (the current period's *entrance* value).

Dataset


KM_estimate
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0
1000
2000
3000
4000
5000
timeline