

Logistisk regression

Georg

5/11/2022

Udforsk betydningen af outcomes:

Det vi gerne vil nu, er at undersøge betydningen af pulsstrykvariationen på patienters 30 dages mortalitet. Derfor skal vi lave en logistisk regression der ser sådan her ud

$$mortality_i \sim PPV_thresh_i, art_mbp_i, Age_i, Sex_i, Asa_i, Eme_ope_i, BMI_i$$

Derfor skal vi bruge et datasæt der ser sådan her ud:

```
Case <- c(29,600,1009,1900)
mortality <- c(0,1,1,0)
PPV_thresh <- c(8,9,6,10)
ART_MBP <- c(54,65,69,70)
Age <- c(22,47,87,4)
Sex <- c(1,0,0,1)
Asa <- c(6,4,1,4)
Eme_ope <- c(1,0,0,1) #emergency operation
BMI <- c(32,16,29,15)
data.frame(Case, mortality,PPV_thresh, ART_MBP, Age, Sex, Asa, Eme_ope, BMI)
```

##	Case	mortality	PPV_thresh	ART_MBP	Age	Sex	Asa	Eme_ope	BMI
## 1	29	0	8	54	22	1	6	1	32
## 2	600	1	9	65	47	0	4	0	16
## 3	1009	1	6	69	87	0	1	0	29
## 4	1900	0	10	70	4	1	4	1	15

Forklaringer:

Case:

- Case skal sorteres således at vi kun finder de operationer der er af typen: “open”, “general surgery” og “general Anesthesia”. Dette kræver en udvidet load funktion. Der burde være lidt under 3000 observationer

PPV_thresh:

- Hvis det så er 180 minutter skal vi fitte 180 gams og udtrække de 180 pulstryk variationer (som er den der amplitude/gennemsnit for pulstryk). Vi vil gerne vide hvor meget af tiden PPV ligger uden for intervallet (6,12) (domain knowledge). Der er lidt forskellige måder dette kan laves på. Men det skal ende ud i enten et samlet tal, eller måske to kolonner med andel af PPV_high, PPV_low.
- Vi kan dnytte “anestart” og “aneend” fra clinical information API'en til kun at vælge det interval af operationen hvor patienten er i narkose.

ART_MBP (lav prioritet) :

- ART_MBP (mean arterial pressure) er faktisk en del af SOLAR8000, så måske skal vi også begrænse vores cases til det?
- Vi er interesseret i hvor ofte ART_MBP ligger under 65 (domain knowledge), så vi skal bare finde andelen af værdierne i ART_MBP under 65

Mortality, Age, sex, asa, Eme_ops og BMI:

- Disse punkter er vores y-variable og nogle confounders vi skal have med, de ligger i clinical information API'en

Konkrete opgaver (prioriteret):

1. Opgrader load funktion så den kan loade ud fra kriterier i clinical information API
 - Høj prioritet. Tror jeg (Georg) tager den da jeg har skrevet det meste af load funktionaliteten.
2. Lav en funktion der tager et langt x antal minutter og outputter x PPV-værdier
 - Høj prioritet
3. Lav funktion, der får et case ID og henter: Age, sex, asa, Eme_ops og BMI
 - (Mellem prioritet) Dette kræver ikke at opgave 1 er løst, så den kan en af jer bare tage
4. Udvid subset funktionen så man kan passe at den kun skal kigge på intervallet mellem "anestart" og "aneend". '
 - (lav prioritet) Kræver måske at opgave 1 er løst, men ikke sikkert

Generelt til udvikling:

- Fra nu af uploader vi ikke flere funktioner til pakken, men vi holder dem i de dokumenter vi skriver i, så samler vi til sidst, så slipper vi for konstant at skulle pushe til master
- Ville være bedst hvis vi stoppede med at pushe til master, men istedet havde hver vores branch vi arbejdede i og så mergede.
- Lad os holde vignetten som den er nu, og så ellers lave et dokument for hver funktion man arbejder på, hvor man kopierer de ting man skal bruge fra vignetten ind.

Spørgsmål:

- Hvis vi skal bruge ART_MBP skal vi vel også restricte vores cases til dem der er optaget på SOLAR8000?
- Skal vi definere PPV ud fra et tal eller to tal som beskrevet i beskrivelsen af PPV_thresh