1. Ambito medicale: Dataset molto semplice e vengono fuori subito cose belle
2. Ambito medicale: dataset difficile e non hanno tanti risultati se non usando tecniche che non ho capito
3. Strade africane: bha
4. Morti: non mi ha colpito particolarmente ma non avevano tante cose
5. Qualcosa con fontana: ci stava, carino
6. Mobilità atene
7. Mobilita dopo covid: sa parlare molto bene, carino ma di fatto facile argomento
8. tosto con img, hanno dovuto cercarsi pacchetti per gestirle. k clustering già fatto circa
9. video fico perchè hanno dati “facili” per renderlo. non ho capito lo scopo, clustering vehicles?
10. viaggi: presentazione bella, dataset facile . groupped features in countries perchè avevano troppi dati. non hanno fatto tanta roba
11. endocardico come il primo: non mi sembra ci abbiano lavorato troppo però hanno detto che hanno un botto di problemi con il dataset mentri quelli del primo gruppo dicevano che era bello quindi bho
12. space: dataset non facile ma argomento noioso: lo scopo è classificare le rocce….
13. space: hanno sfruttato la presentazione di quelli di prima hahaha che grandi. troppo trash, hanno esagerato tra meme e foto di secchi…. non hanno detto tanto come quelli di prima. devono aspettare la fine del corso
14. financial: understanding financial mkt, ha la voce da morto, c’è un cinese ihihihi, minchia che noia. questi non hanno fatto un cazzo e stanno cercando di occupare i 5 min. no vabbhe sparatemi stanno pure superando i 5 min con sta noia e questa cadenza
15. ambito medicale: come il 2o gruppo ma sembra che abbiano dei risultati più belli anche se hanno fatto vedere molti meno grafici quindi forse non hanno detto tutto come i secondi
16. spiegato bene pb. obiettivi chiari
17. img tumori, casuality between patological data and … bonferroni e pca accennati. direi dataset medico facile e sembra abbiano fatto tante cose (tanti problemi e idee chiare su come risolverli)
18. (hanno fatto dopo per pb tecnici) culture institutions; noiosi. stat cose, hanno fatto box cox per assunzione gaussianità ok bravi ma ci stanno troppo (anche secchi sta per addormentarsi)
19. mug eruption tipo. grouped outliers e guardano l’andamento di questi (time series of outliers). ma poco chiaro inizio (seguiti da mara …). ggplot2 package for plots carini
20. spotify interessante sull’algoritmo. ora funziona su deep neural network ma loro vogliono fare cose “più classiche” in alternativa. tanti data, detto obiettivi no analisi ancora fatte. tanto preprocessing in pyhton
21. changing movements before and after covid. non ho capito che info hanno. pca sembra interessante. work on trajectories so they interpolate them
22. presentazione fica graficamente. seri ma tengono anche attenzione con gli alpaca, genius. reduced dim of their data. **mappa europa con barplot** -> **kepler** **https://kepler.gl/**
23. saltato
24. human mobility: sembra ben fatto, dataset facile
25. fichisssimoo
26. 1
27. 2
28. 3
29. 4
30. look if an area is attractive by looking at flow of people. log di stops, distance home wrt dwell. pc1=mean of day pc2= week vs weekend.
31. bankrupt dataset. tolto outliers. pca prima e dopo la trasformazione sulle features. naive classifier ma risultati schifosi. aim: predict bankrupt with a logistic regression.
32. moralità in italia. hanno guardato il fumo a milano nelle varie stagioni. aim: guardare maschi vs femmine e fare geographical clustering.
33. climate change dal satellite. 4 dataset usano (comparano 2 alla volta). aim: test x misure ripetute + regression sul livello dell’acqua.

COMMENTI SECCHI: **quelle anche per noi magari**

2) occhio a pca, meglio proiettare i dati in uno spazio 2d o 3d piuttosto che usare greedy algorithms e non saper dare interpretazioni

3) se avere priors usatele per fare clustering

**5) clear. references to the tutor and faces**

4) guardare il tasso di mortalità se è correlato con situazione socio economica

6) c**onnections between spatial domain -> spatial statistics.** you could try to cluster in order to find some characteristics

7) vedere i cambiamenti nel tempo guardando prima e dopo.

8) condividere info con l’altro gruppo. spatial autocorrelation sarà utile

9) local diffusion tensor (matrix sym def pos) describing movements and so some spatial analysis. mara has ideas for sure. (ma serve functional data analysis?)

10) rescale wrt densities of people (cina 1 miliardo vs italia 60 milioni). compositional data -> distances between distributions to see some changes. aggregate countries in europe vs asia maybe.

11) hp not well stated. H0: mean = sample mean (not useful) (say mean of population). clear slides.

12) alignment to take into account different behaviours of temperatures.

13) smoothing to identify the min and max of temperatures. use fourier (functional data analysis if data is periodic) or something to identify max and min.

14) se riesci sei un figo

15) try to use multidimensional scaling and iso maps. goals are interesting (find a way to mimic data with easier tools)

16) on what variable you want to apply clustering

17) clear problems and how to solve them. pochi dati, problem of identifying outliers. prior info could be some thoughts about the hospital where to be treated? say always the meaning of the model you use (random stat model for instance)

18) ops. li ha corretti su using “traslating”, cioè hanno fatto translation dopo box cox o prima, hanno fatto confusione su sequenza di cambiamenti dei dati credo. non usare il p value per capire se H0 è buono (come ipotesi), ma dire se H0 si rifiuta o no. prima fare transformation con boxcox ma si possono trovare trasformazioni migliori.

19) smooth the data and treat it as spatial data. use some exploratory tool for spatial autocorrelation (Lism) in order to find hotspots.

20) say how you checked the prediction (CV, test set ecc).

21) questo dataset dà già info of the “new normal” or we are coming back to before covid? ci sono percorsi che rimangono tali per cui attenzione a questa cosa (**anche noi?)**

22) find attractors to understand better the problem. images from alpacas and kepler (?) jacopo ghirri da sentire

23) distance matrix that could change in time. see also clusters changing of time. maybe standardize wrt residences.

24)

25) che fighi sono quelli? ha detto di portare avanti in parallelo entrambi gli approcci cioè cbgs come punti e come aree(densità e no, chiaramente cambierebbe la interpretazione). poi da capire bene lo scopo sotto clustering -> lui suggerisce clustering di features? non so se ho capito bene

26)

….

29) say where data come from.

30) non avete un cazzo, speriamo che a luglio possiate mostrare qualcosa.

31) distributions concentrated to a point. variables are discrete or the support very small wrt others? 3 pca is fine, no more. which is the advantage of going to an higher dim space? naive classifier: change the threshold to look at fp and fn (take into account the cost of misclassification).

32) settimana prossima secchi pubblica una ricerca sul spatial dependence mortality. sapendo che certe features provocano mortalità si può esaltare sta cosa

33) focalizzarsi su un argomento specifico (possono fare spatial data o time data potenzialmente).

PRESENTAZIONE DI LUGLIO

presentazione velocissima dei dati

problema

soluzioni

risultati