

پرسش های کلاسی

سری شماره یک

پرسش ۱

مفاهیم Big Data و Machine learning و Artificial intelligence و Data Mining و Data Science و NLP و Text mining را تعریف کنید و شرح دهید و همچنین تفاوت ها و اشتراکات این حوزه ها را نیز به همراه مثال بررسی کنید . (حداقل دو صفحه A4)

پرسش ۲

- مجموعه داده زیر را که مربوط به بیماران پارکینسون است از نظر ساختار و جنس داده و ویژگی ها بررسی کنید و در مورد آن توضیح دهید. این مجموعه داده را با استفاده از "Pandas" تبدیل به یک جدول داده کنید و توضیح دهید که چگونه می توانید برچسب را انتخاب کنید. برچسب را در یک متغیر و داده (نمونه) ها را در متغیری جداگانه (جدول) ذخیره کنید. در این صورت داده ها و ویژگی ها به چه تعداد هستند؟ سپس حداقل سه مقاله از ۲۰۱۸ به بعد بیابید که بر روی این مجموعه داده کلاس بندی انجام داده اند. بدون ذکر جزئیات توضیح دهید که روش آنها چه بوده است و به چه دقتی دست یافته اند.

Link: <https://archive.ics.uci.edu/ml/datasets/parkinsons>

- با استفاده از "NetworkX" در پایتون برنامه ای بنویسید که داده را به صورت بصری رسم کند. (برچسب ها در خروجی دیده شود) . با کمک این پکیج تعداد یالها و راس ها را نیز گزارش کنید. گراف را طوری رسم کنید که اندازه بصری هر راس با اندازه درجه آن راس متناسب باشد. از فایل های edge_idx و node_lbls که مربوط به یال ها و برچسب هر گره هستند استفاده کنید تا به اطلاعات داده برسید. سپس اندازه بصری گره های گراف را طوری رسم کنید که صرفا تعداد همسایه هایی که از کلاس مشابه آن گره هستند در اندازه تاثیر گذار باشد.
- مجموعه داده های زیر را در نظر بگیرید و به سوالات زیر پاسخ دهید.

- نوع داده چیست؟ تعداد نمونه ، ویژگی و ... به چه تعداد و چه ابعادی است؟ در مورد کاربرد این نوع از داده ها تحقیق کنید.
- با پایتون این داده را فراخوانی کنید .
- تصویر یکی از کانال ها را رسم کنید.
- برای چند نقطه مختلف دیاگرام بردار اسپکتروم (Spectrum Plot)(مقدار هر یک از نقاط در باندهای مختلف) در یک نمودار رسم کنید. (مقادیر بین ۰ تا ۲۵۵ نرمال سازی شوند).
- برچسب این داده را نیز با پایتون بخوانید و نمایش دهید.

Link: <https://rslab.ut.ac.ir/data>

— Pavia University scence —

پرسش ۳

در طراحی آزمون محک (Benchmark) "NIPS 2003" از ۵ مجموعه داده استفاده شده است که از طریق لینک زیر قابل مشاهده است (در صورت نیاز به اطلاعات بیشتر وظیفه جست و جو با دانشجو می باشد.)

- تحقیق کنید این چالش چه هدفی داشته است و مدل های شرکت کننده های در این چالش چگونه و با چه معیارها و چه روش های اندازه گیری ای مورد ارزیابی و رتبه بندی قرار گرفته اند و مقالات برنده در بخش های مختلف چه روشی داشته اند. (نیاز به خواندن مقاله های مدل های شرکت کننده در مسابقه نیست و صرفا نوع مدل ذکر شود اما برای توضیحات خودآزمون محک حتما مقاله مرتبط به آن خوانده شود).
- تحقیق کنید که این مجموعه داده ها هرکدام چند داده و چند ویژگی دارند. از منظر تنک بودن (Sparsity) ، در چه وضعیتی هستند و طراحی آنها به طور خلاصه چه مراحلی را طی کرده است و نحوه تقسیم بندی train/val/test به چه روشی صورت گرفته است.
- بررسی کنید که دلیل استفاده از این مجموعه داده ها (از نظر جنس و وضعیت داده ها) در این چالش چه بوده است؟
- برای پیش پرداخت (Pre-processing) این داده ها قبل از استفاده چه روش هایی در این آزمون معرفی شده است؟ (مقاله طراحی این چالش را مطالعه کنید).

Link: <http://clopinet.com/isabelle/Projects/NIPS2003>

Datasets: ARCENE, GISETTE, DEXTER, DOROTHEA, MADELON

دوره یادگیری ماشین / دانشگاه تربیت مدرس / گروه علوم کامپیوتر / نیمسال اول سال تحصیلی ۱۴۰۱-۱۴۰۲
استاد درس : دکتر منصور رزقی آهق (Mrezghi.ir)