



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

نامپای

۱. موارد خواسته شده را «بدون کتابخانه» (بدون استفاده از توابع داخلی، به کمک «map و lambda» و با «list comprehension») و با «نامپای» پیاده سازی کنید. (۱ نمره)
 ۱. برداری با ۱۰۰۰ عدد تصادفی از بین اعداد ۱۰ تا ۱۰۰۰ ایجاد کنید.
 ۲. میانگین و واریانس این اعداد را محاسبه کنید.
۲. موارد خواسته شده را «بدون کتابخانه» (بدون استفاده از توابع داخلی، به کمک «map و lambda» و با «list comprehension») و با «نامپای» پیاده سازی کنید. (۱ نمره)
 ۱. یک ماتریس ۱۰ در ۱۰ دلخواه ایجاد کنید.
 ۲. مقادیر این ماتریس را به صورت ستونی و ردیفی جمع کنید.
۳. موارد خواسته شده را به کمک نامپای پیاده سازی کنید. (۴ نمره)
 ۱. با استفاده از تابع `np.array()` آرایه «heights» بسازید تا ارتفاع های ۵ گونه مختلف گیاه را (به سانتیمتر) نمایش دهد. از مقادیر زیر استفاده کنید: [۱۵۰، ۱۲۰، ۱۷۵، ۲۰۰، ۱۴۰]، همچنین مطمئن شوید که نوع داده آرایه `np.float64` باشد.
 ۲. با استفاده از تابع `np.array()` یک آرایه دیگر «widths» بسازید تا عرض های همان ۵ گونه گیاه را (به سانتیمتر) نمایش دهد. از مقادیر زیر استفاده کنید: [۷، ۱۰، ۱۵، ۸، ۱۲]، همچنین مطمئن شوید که این آرایه دارای نوع داده `np.float64` باشد.
 ۳. مساحت هر گونه گیاه را با ضرب المان های متناظر آرایه های `heights` و `widths` محاسبه کنید و نتیجه را در یک آرایه نامپای «areas» ذخیره کنید.
 ۴. با استفاده از تابع `np.arange()` یک آرایه «indices» بسازید تا یک آرایه شاخص را از ۰ تا تعداد گونه های گیاه (شامل خود عدد نمی شود) ایجاد کند. (این به عنوان یک شاخص برای گونه های گیاه استفاده خواهد شد)
 ۵. با استفاده از تابع `np.array()` آرایه دوبعدی «plant_info» را بسازید. ستون اول باید شاخص های گیاه را نمایش دهد، ستون دوم باید ارتفاع ها را نمایش دهد و ستون سوم باید عرض ها را نمایش دهد. شاخص ها، ارتفاع ها و عرض ها را در این آرایه دوبعدی ترکیب کنید.
 ۶. حالا با استفاده از تابع `np.eye()` یک ماتریس همانی ۳x۳ ایجاد کنید و به عنوان «identity_matrix» ذخیره کنید.
 ۷. با استفاده از تابع `np.diag()` یک ماتریس قطری «diagonal_matrix» ایجاد کنید. المان های قطر این ماتریس باید با مساحت گونه های گیاه ذخیره شده در آرایه مساحت ها مطابقت داشته باشند.
 ۸. در نهایت، از تابع `np.fromfunction()` برای ایجاد آرایه «squared_sums» استفاده کنید. این آرایه باید حاوی مجموع مربعات شاخص ها و ستون ها باشد. از لامبدا برای تعریف مقادیر در این آرایه استفاده کنید.



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

۴. تصور کنید که ابعاد آرایه «x» 4 در 5 است و نقشه‌ی یک گنج مخفی را نمایش می‌دهد. هر سلول در این آرایه مقداری را که عمق در آن مکان قرار دارد را نشان می‌دهد. هدف شما پیدا کردن گنج است، که در یک مکان با عمق کمتر یا مساوی ۲ مخفی شده است. مراحل زیر را برای پیدا کردن گنج باید دنبال کنید: (۴ نمره)
۱. ابعاد، اندازه و نوع داده آرایه «x» را محاسبه کنید.
 ۲. آرایه «x» را به شکل یک آرایه با ابعاد 2 در 1- تغییر شکل دهید.
 ۳. کل ردیف اول آرایه «x» را نمایش دهید.
 ۴. کل ستون اول آرایه «x» را نمایش دهید.
 ۵. کل ردیف آخر آرایه «x» را نمایش دهید.
 ۶. عمق مکان (1,1) در نقشه چیست؟
 ۷. ستون‌های اندیش ۱ تا ۳ (شامل خود اعداد) را نمایش دهید.
 ۸. ماسک Boolean برای خانه‌هایی که عمق آن‌ها بیش‌تر یا مساوی ۲ است تعریف کنید.
 ۹. به کمک ماسک تعریف شده تمامی خانه‌هایی که عمق آن‌ها بیشتر از ۲ است را پیدا کنید.
 ۱۰. عمق خانه‌ها با موقعیت (0,1) و (0,2) را نشان دهید.
۵. موارد خواسته‌شده را به کمک نامپای پیاده‌سازی کنید. (۶ نمره)
۱. **ایجاد آرایه:** برنامه‌ای بنویسید تا یک آرایه NumPy ایجاد کند که شامل ۱۰۰ عدد بین ۰ و ۱ با فاصله مساوی باشد.
 ۲. **عملیات آرایه:** یک تابع پیاده‌سازی کنید که شباهت کسینوسی بین دو آرایه NumPy (که نشان دهنده دو بردار است) را محاسبه کند. چگونه از این تابع برای پیدا کردن سندهای مشابه در یک مجموعه متنی استفاده می‌کنید؟
 ۳. **اندیس‌گذاری منطقی:** یک آرایه NumPy را ایجاد کنید که شامل سن گروهی از افراد باشد و با استفاده از اندیس‌گذاری منطقی، افرادی را شناسایی کنید که بین ۱۸ و ۳۰ سال سن دارند.
 ۴. **تکرار آرایه:** یک مقایسه بین حلقه for و رویکرد برداری شده برای محاسبه مجموع تجمعی عناصر در یک آرایه NumPy بزرگ بنویسید. تفاوت عملکرد را با استفاده از تابع timeit نشان دهید.
 ۵. **ماسک و فیلتر آرایه:** با داشتن یک آرایه NumPy که شدت‌های پیکسل را در یک تصویر نشان می‌دهد، کدی بنویسید که یک ماسک دودویی ایجاد کند که تمام پیکسل‌هایی با شدت بیشتر از حداقلی مشخص را برجسته کند.
 ۶. **عملیات پیشرفته:** از تابع SVD NumPy برای کاهش ابعاد در یک مجموعه داده استفاده کنید. آن را برای کاهش ابعاد مجموعه داده از ۱۰۰ به ۱۰ و در عین حال حفظ بیشترین واریانس، اعمال کنید.



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

۷. **توابع آماری:** تابعی پیاده‌سازی کنید که ضریب همبستگی پیرسون بین دو آرایه NumPy که نماینده سری‌های داده هستند را محاسبه کند. این تابع را بر روی یک مجموعه داده واقعی اعمال کنید تا همبستگی‌ها را پیدا کنید.

۸. **ورودی/خروجی فایل:** یک برنامه بسازید که یک فایل CSV حاوی داده‌های عددی را بخواند و آن را به یک آرایه NumPy تبدیل کند. آرایه حاصل را در یک فایل دودویی ذخیره کنید.

۹. **پردازش سیگنال:** یک برنامه پایتون بنویسید که یک سیگنال صوتی را به عنوان یک آرایه numpy دریافت کرده و تبدیل فوریه را بر روی آن اعمال کرده و اجزای فرکانسی آن را تجزیه و تحلیل کند. طیف فرکانس را نمایش دهید.

۶. شما در حال تجزیه و تحلیل داده‌هایی هستید که از حسگرهای مختلف جمع‌آوری شده‌اند. این داده‌ها در یک آرایه NumPy به نام «sensor_data» ذخیره شده‌اند که ردیف‌ها نماینده حسگرهای مختلف و ستون‌ها نماینده نقاط زمانی مختلف هستند. هدف شما انجام یک سری مراحل پیش‌پردازش داده‌ها است تا آماده‌سازی آنها برای تجزیه و تحلیل بیشتر انجام شود. (۵ نمره)

داده‌های اولیه:

```
sensor_data = np.array([[1.2, 3.5, 2.4, 1.8],  
                        [4.1, 2.2, 3.9, 4.8],  
                        [2.8, 1.9, 4.5, 3.2]])
```

۱. **استانداردسازی بر اساس میانگین و انحراف معیار ستونی:** استانداردسازی را بر روی آرایه «sensor_data» انجام دهید، با کم کردن میانگین هر ستون و سپس تقسیم بر انحراف معیار هر ستون. نتیجه را در یک آرایه جدید به نام «standardized_data» ذخیره کنید.

۲. **استانداردسازی بر اساس میانگین و انحراف معیار ردیفی:** استانداردسازی را بر روی آرایه «sensor_data» انجام دهید، با کم کردن میانگین هر ردیف و سپس تقسیم بر انحراف معیار هر ردیف. نتیجه را در یک آرایه جدید به نام «row_standardized_data» ذخیره کنید.

۳. **محاسبه مجموع تمام عناصر:** مجموع تمام عناصر در آرایه «sensor_data» را پیدا کنید.

۴. **محاسبه مجموع ستون‌ها:** مجموع هر ستون در آرایه «sensor_data» را محاسبه کنید.

۵. **محاسبه مجموع ردیف‌ها:** مجموع هر ردیف در آرایه «sensor_data» را محاسبه کنید.

۶. **ترکیب آرایه‌ها:** آرایه «sensor_data» و آرایه «standardized_data» را افقی (کنار هم) در یک آرایه جدید به نام «combined_data» ترکیب کنید.



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

۷. تقسیم آرایه‌ها: آرایه «combined_data» را به سه آرایه تقسیم کنید: A، B و C. به گونه‌ای که هر آرایه شامل یک سوم ستون‌ها باشد.
۸. محاسبه ضرب ماتریسی: ضرب ماتریسی بین آرایه «sensor_data» و ترانهاده آن (sensor_data.T) را انجام دهید و نتیجه را در یک آرایه جدید به نام «matrix_product» ذخیره کنید.
۹. محاسبه عملیات عنصر به عنصر: عملیات عنصر به عنصر زیر را محاسبه کنید و هر نتیجه را در یک آرایه جداگانه ذخیره کنید:
- «log_data»: لگاریتم طبیعی (با پایه e) آرایه «sensor_data».
 - «exp_data»: نمایش از آرایه «sensor_data».
۱۰. ترکیب و مجدداً تقسیم: آرایه «sensor_data» را با خود به صورت عمودی ترکیب کنید (یکی را بالای دیگری قرار دهید) در یک آرایه جدید به نام «stacked_data». سپس، آرایه «sensor_data» را عمودی به دو آرایه، «split_1» و «split_2» تقسیم کنید، به گونه‌ای که آرایه اول شامل نیمه اول ردیف‌ها و آرایه دوم شامل نیمه دوم ردیف‌ها باشد.
۷. موارد خواسته شده را «بدون کتابخانه» (بدون استفاده از توابع داخلی، به کمک «map و lambda» و با «list comprehension») و با «نامپای» پیاده‌سازی کنید. (۴ نمره)
۱. ۱۰ نقطه‌ی تصادفی ۲۰ بعدی ایجاد کنید.
۲. برای هر نقطه، نزدیک‌ترین و دورترین نقطه موجود از بین نقاط دیگر را با کمک فاصله اقلیدسی پیدا کرده و اندیس و نقاط مربوطه را گزارش کنید.
۸. شما داده‌ها را از سه آزمایش مختلف جمع‌آوری کرده‌اید، هر یک از آنها یک آرایه اندازه‌گیری‌ها ایجاد کرده‌اند. هدف شما از ترکیب و تحلیل این آرایه‌ها، استخراج برداشت‌های معنی‌دار است. (۴ نمره)
- آماده‌سازی داده: ابتدا، داده‌های آزمایشی را به آرایه‌های NumPy بارگذاری می‌کنیم:



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

Data from Experiment 1

```
exp1_data = np.array([[1, 2, 3],  
                      [4, 5, 6],  
                      [7, 8, 9]])
```

Data from Experiment 2

```
exp2_data = np.array([[10, 11, 12],  
                      [13, 14, 15],  
                      [16, 17, 18]])
```

Data from Experiment 3

```
exp3_data = np.array([[19, 20, 21],  
                      [22, 23, 24],  
                      [25, 26, 27]])
```

۱. **ترکیب آرایه ها:** اولین وظیفه شما ترکیب سه آرایه آزمایشی به یک آرایه تکی است. با این وجود، قبل از انجام

این کار، نیاز دارید که هر سه آرایه را به صورت مستقل استانداردسازی کنید. استانداردسازی شامل کم کردن میانگین و تقسیم بر انحراف معیار است. شما باید به صورت ستونی استانداردسازی کنید (به این معنا که هر ستون پس از استانداردسازی میانگین ۰ و انحراف معیار ۱ را داشته باشد).

۲. **ترکیب و تقسیم مجدد آرایه ها:** بعد از آن، شما باید آرایه های استانداردسازی شده از وظیفه ۱ را به یک آرایه

ترکیب کنید. سپس، این آرایه ترکیب شده را به دو آرایه جدید تقسیم کنید: یکی شامل دو ردیف اول از آرایه ترکیب شده و دیگری شامل ردیف آخر باشد.

۳. **محاسبه ضرب داخلی:** حالا دو آرایه از وظیفه ۲ دارید. محصول داخلی بین این دو آرایه را محاسبه کنید.

۴. **انجام جمع انیشتین:** در نهایت، روی نتیجه وظیفه ۳ یک عمل جمع انیشتین انجام دهید. این جمع باید یکی از ابعاد را کوچک کند، به گونه ای که در نهایت به یک عدد اسکالر (آرایه صفر بعدی) برسید.

۹. داده های جمعیتی چند ساله برای سه گونه به شما ارائه شده است: هویج، خرگوش، و سیاه گوش. با این حال، برخی از داده ها برای سال های خاصی گم شده یا خراب شده اند. «سال های بد» خاص که نمی توان از داده ها برای محاسبات استفاده کرد، به شرح زیر است: (۴ نمره)

سال های نامطلوب؛ ۱۹۰۰، ۱۹۰۱، ۱۹۰۲، ۱۹۰۳، ۱۹۰۴، ۱۹۱۳، ۱۹۱۵، ۱۹۲۲، ۱۹۲۴، ۱۹۴۵ و ۱۹۴۶

برای انجام تحلیل های خود، باید از NumPy استفاده کنید. اگرچه داده های ۱۲۰ ساله از ۱۹۰۰ تا ۲۰۲۰ را در اختیار دارید، تجزیه و تحلیل باید بر ۵۰ سال اول متمرکز شود، به استثنای «سال های نامطلوب» که در بالا ذکر شده است.

وظیفه شما ارائه موارد زیر است:



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی
امیرمحمد کوبش پور

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

دکتر سامان هراتی زاده

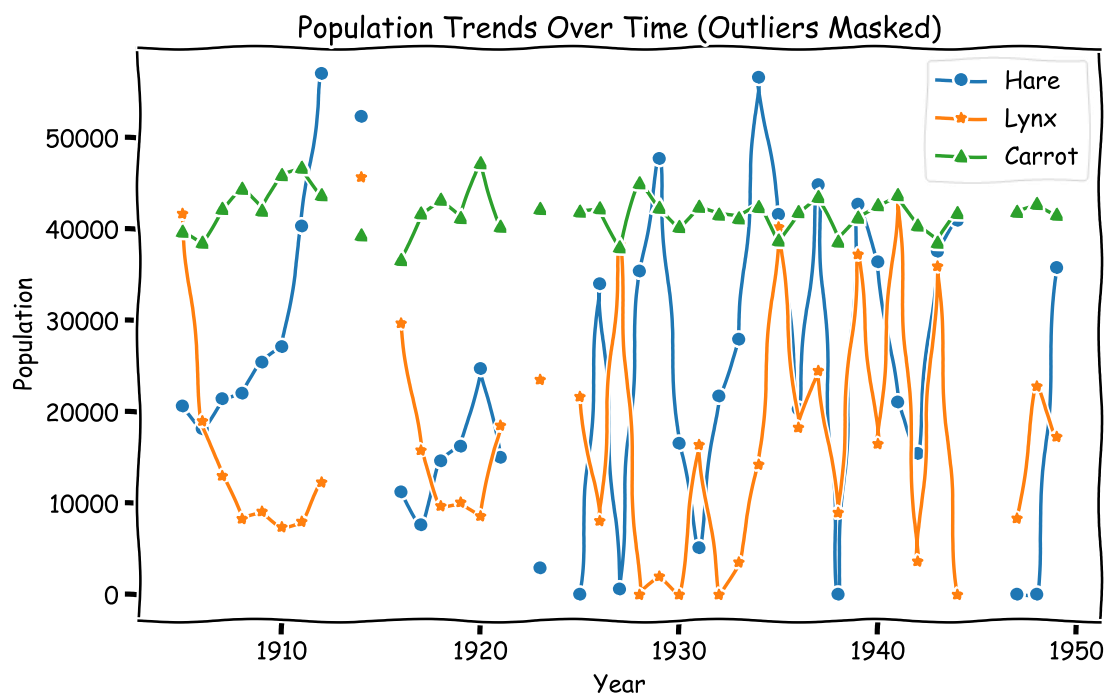
دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

انحراف معیار و میانگین جمعیت خرگوش‌ها، سیاه‌گوش‌ها و هویج‌ها را برای ۵۰ سال اول به استثنای «سال‌های نامطلوب» محاسبه کنید.

نموداری ایجاد کنید که روند جمعیت را در طول زمان برای خرگوش‌ها، سیاه‌گوش‌ها و هویج نشان می‌دهد، داده‌های «سال‌های نامطلوب» پوشانده یا حذف شوند.

لطفاً این مراحل را اجرا کنید و گزارش مفصلی از یافته‌های خود ارائه دهید.



شکل ۱: تصویر نهایی مورد انتظار از پیاده‌سازی شما

۱۰. یکی از دیگر فوت و فن‌های NumPy، ساختن یک ماتریس leave-one-out از یک بردار با طول k است. در این ماتریس، هر سطر یک بردار با طول $k-1$ است که هر بار یک عضو مختلف حذف می‌شود. از این روش برای ارزیابی دقت مدل پیش‌بینی کننده بر دادگان دیده نشده استفاده می‌شود. به عنوان مثال، فرض کنید نقاط داده $[(1, 4), (2, 7), (3, 11), (4, 9), (5, 15)]$ را می‌خواهید برای یک مدل رگرسیون ساده با روش LOOCV مورد آزمایش قرار دهید. برای هر مرحله اعتبارسنجی متقابل، یک نقطه را برای آزمایش استفاده کرده و 4 نقطه دیگر را برای آموزش استفاده می‌کنید. به عبارت دیگر، می‌خواهید مجموعه آموزش به شکل زیر باشد: (۶ نمره)



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

[(2, 7), (3, 11), (4, 9), (5, 15)]
[(1, 4), (3, 11), (4, 9), (5, 15)]
[(1, 4), (2, 7), (4, 9), (5, 15)]
[(1, 4), (2, 7), (3, 11), (5, 15)]
[(1, 4), (2, 7), (3, 11), (4, 9)]

- با در نظر گرفتن داده اولیه و خروجی‌های حاصل، پیاده‌سازی ارائه کنید که نتایج تولیدی نمونه را حاصل می‌شوند.
۱۱. مجموعه‌ای از دادگان گل‌های زنبق در مسیر `data/iris` قرار داده شده است، سلولی برای بارگذاری دادگان در نوت‌بوک قرار دارد. در این تمرین قصد داریم فاصله اقلیدسی^۱ مجموعه گل‌های جدید را از مجموعه موجود بسنجیم. فاصله اقلیدسی یکی از روش‌های محاسبه فاصله نمونه‌ها در فضای داده از هم می‌باشد. (۴ نمره)
۱. فاصله را با کمک دو حلقه محاسبه کنید.
 ۲. فاصله را با کمک یک حلقه محاسبه کنید.
 ۳. فاصله را بدون حلقه و فقط با `vectorization` محاسبه کنید.
 ۴. برای تمام موارد پیشین به کمک `timeit` زمان اجرا را گزارش کنید.
۱۲. با توجه به دادگان بارگذاری شده در سلول موجود در نوت‌بوک ارائه شده، به سوالات زیر پاسخ دهید و آن‌ها را گزارش کنید. (۵ نمره)
۱. کدام دارای بیشترین نرم^۲ (فاصله اقلیدسی از مبدا) است.
 ۲. دو `patent` را پیدا کنید که بیشترین فاصله را از هم دارند.
 ۳. تعداد `patent`هایی که با نزدیک‌ترین همسایه خود در یک دسته^۳ هستند را به دست بیاورید.
۱۳. به شما مجموعه داده `X` با `N` نقطه داده و مجموعه داده دیگر `Y` با `M` نقطه داده، در `Rd` داده شده است. تابعی بنویسید که موارد زیر را انجام دهد: (۹ نمره)
۱. مجموعه داده‌های `X`، `Y` و معیار فاصله (مانند «Euclidean» یا «cosine») را به عنوان پارامترهای ورودی قبول کند.
 ۲. برای محاسبه ماتریس کامل فاصله (به نام `D` با اندازه‌ی `N*M`) بین `X` و `Y` با استفاده از معیار فاصله مشخص شده، از `scipy.spatial.distance.cdist` استفاده کند.
 ۳. از `scipy.spatial.distance.pdist` برای محاسبه ماتریس فاصله متراکم `S` برای مجموعه داده `X` استفاده کند.
 ۴. ماتریس `S` را به ماتریس فاصله مربعی مجانب `R` (`redundant square distance matrix`) با استفاده از `scipy.spatial.distance.squareform` تبدیل کند.

¹ [Euclidian Distance](#)

² Norm

³ category



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

[امیرمحمد کوشش پور](#)

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

۵. در نهایت تابع نوشته شده ماتریس D ، R و کل تعداد محاسبات فاصله انجام شده را برگرداند.

در گزارش خود موارد خواسته شده را ارائه کنید.

- چگونه می‌توانید از محاسبات اضافی فاصله در تبدیل S به R اجتناب کنید؟
- استفاده از ماتریس فاصله متراکم S چگونه تعداد محاسبات فاصله را کاهش می‌دهد؟
- پیچیدگی اجرایی پیاده‌سازی شما به ازای N ، M و d چیست؟
- پیاده‌سازی شما باید چه تغییری کند اگر X یا Y دارای مقادیر گمشده/NaN باشند؟
- در مورد مزایا و معایب استفاده از ماتریس‌های فاصله متراکم در تمرین بحث کنید - چه زمانی بهتر است فقط از ماتریس کامل $N \times M$ استفاده کنید؟
- راه‌هایی برای بهبود بیشتر کارایی پیشنهاد کنید - سایر ساختارها/الگوریتم‌های داده‌سازی می‌توانند بر مبنای ماتریس‌های فاصله تراکم شده بنا شوند را بیان کنید.

مجموعه داده‌های X و Y که تابع شما دریافت خواهد کرد، به صورت زیر خواهد بود:

- X یک آرایه numpy با اندازه $N \times D$ است که شامل N نقطه داده است.
- Y یک آرایه numpy با اندازه $M \times D$ است که شامل M نقطه داده است.
- در اینجا D تعداد ابعاد داده‌ها (تعداد ویژگی‌ها) است.
- تابع شما نیازی به تولید داده ندارد - آرایه‌های X و Y به عنوان ورودی دریافت می‌شوند.
- می‌توانید فرض کنید که N ، M و D تماماً اعداد صحیح مثبت هستند.
- پیاده‌سازی شما باید برای هر مجموعه داده ارائه شده X و Y با این خصوصیات کار کند، تست داده‌های مختلف را برای پیاده‌سازی خود ارائه کنید.



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

پانداس

۱. مجموعه‌ای از اسامی غذاها در فایل food.csv در مسیر data/food قرار داده شده است. دادگان را به کمک پاندا به صورت دیتافریم بخوانید (توضیحات ستون‌های مجموعه‌دادگان در نوت‌بوک آورده شده است). (۱ نمره)
 ۱. می‌خواهیم بدانیم چه نسبتی از کل غذاها تند مزه هستند؟ (یعنی حاصل تقسیم تعداد غذاهای تند مزه بر تعداد کل غذاها چند است) نتیجه را تحت عنوان spicy_ratio گزارش کنید.
 ۲. تعداد ۱۵ غذایی که بیشترین زمان را مصرف می‌کنند (منظور از زمان، مجموع زمان آماده‌سازی و زمان طبخ مدنظر است) را به ترتیب نزولی در قالب یک دیتافریم در متغیر df_most_time_consuming ذخیره کنید.
۲. در این سوال دادگان شبکه اجتماعی گرام در اختیار شما قرار گرفته است. جهت بارگذاری دادگان سلول مرتبط را اجرا کنید. در این شبکه اجتماعی ۵ نوع طبقه‌بندی وجود دارد. این دسته‌بندی بسته به شرایط و اسم پیج متفاوت خواهد بود. در این تمرین از شما خواسته می‌شود که اعضای دسته‌های مختلف را نشان دهید. (۳ نمره)
 ۱. دسته‌ی برنجی؛ افرادی هستند که توسط پیج دنبال می‌شوند (following) از ۱۱ کمتر باشد.
 ۲. دسته‌ی مسی؛ افرادی هستند که تعداد پست‌های پیجشان از تعداد افرادی که آن پیج دنبالشان می‌کند (following) بیشتر باشد.
 ۳. دسته‌ی برنزی؛ افرادی هستند که تعداد دنبال‌کنندگان پیجشان بیشتر از یک میلیون نفر باشد.
 ۴. دسته‌ی نقره‌ای؛ افرادی که دارای دو شرط زیر را همزمان داشته باشند.
 - تعداد دنبال‌کنندگان پیجشان بیشتر مساوی ده میلیون باشد.
 - تعداد دنبال‌شوندگان توسط پیج کمتر از ۵۰ باشد.
 ۵. دسته‌ی طلایی؛ افرادی که حداقل یکی از دو شرط زیر را داشته باشند (یعنی شرط اول یا دوم)
 ۶. تعداد دنبال‌کنندگان (follower) بزرگتر از ده میلیون و تعداد دنبال‌شوندگان (following) کمتر از ۲۰ نفر باشد.
 ۷. تعداد پست‌ها کمتر از ۱۵ و حرف R در اسم صاحب پیج وجود داشته باشد. (مثل lAzaR)
۳. در این سوال مجموعه‌ای از دادگان سایت goodreads در مسیر data/goodreads قرار گرفته است. همانطور که می‌بینید در دیتاست اصلی (کتاب)، سه متغیر lang، author و publisher تنها به صورت id هستند. (۳ نمره)
 ۱. برای استفاده بهتر از کل اطلاعات، سه دیتابیس دیگر را با توجه به شماره نمایه‌ی آن‌ها به این دیتابیس اضافه کنید.



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

[امیرمحمد کوشش پور](#)

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

۲. حال که تمام مقادیر اضافه شده‌اند، نیازی به سه ستون id که به عنوان کلید استفاده کردیم نیست، پس آن‌ها را حذف می‌کنیم.
۳. نام ستون‌ها را طوری تغییر دهید که جدول نهایی چیزی شبیه به جدول آماده شده در نوت‌بوک باشد.
۴. قصد داریم که تمام کتاب‌های خود را از یک انتشاراتی خریداری کنیم اما برای اینکار نیاز به میانگین امتیاز هر کدام از انتشاراتی‌ها داریم تا بتوانیم بهترین‌ها رو فقط انتخاب کنیم. به این منظور میانگین امتیاز کتاب‌های هر انتشاراتی را حساب کرده و گزارش کنید.
۵. می‌خواهیم بدانیم که هر نویسنده چند کتاب با تعداد صفحات مشخصی نوشته است؛ به این منظور تعداد صفحات کتاب را به صورت گسسته و در بازه‌های کمتر از ۲۵۰، ۲۵۰ تا ۴۹۹، ۵۰۰ تا ۷۴۹، ۷۵۰ تا ۹۹۹ و بزرگتر مساوی ۱۰۰۰ در نظر بگیرید. در نهایت جدول خروجی که مانند جدول زیر خواهد شد را گزارش کنید.

A.B. Yehoshua	250-499	1
	500-749	1
A.D.P. Briggs	250-499	1
A.E. Cunningham	<250	1
A.J. Jacobs	250-499	1



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی
امیرمحمد کوشش پور

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

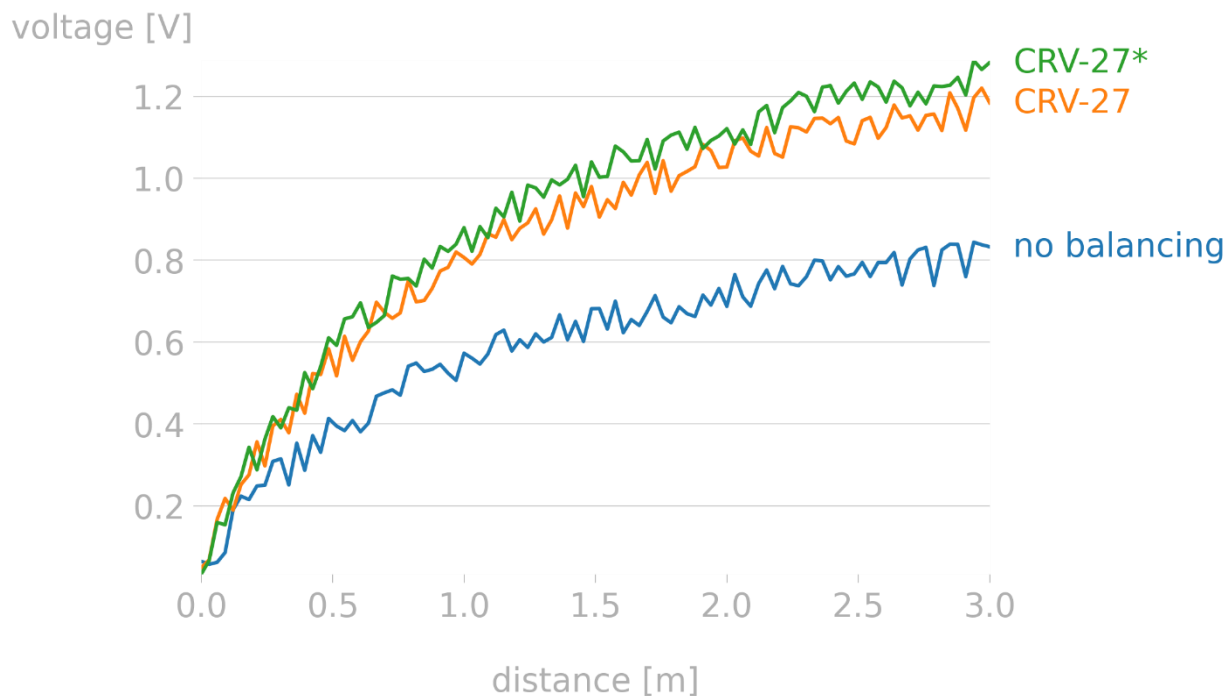
مت پلات لیب^۴

۱. از بین دادگان موجود در مسیر data محتوای فایل فشرده logo.zip را بخوانید. محتوای موجود، لوگوی دانشگاه تهران است. (۱ نمره)

۱. تصویر را ترسیم کنید و ابعاد آن را گزارش کنید. ترسیم خود را نیز در گزارش درج کنید.

۲. برای داده‌های زمانی^۵ بهتر است همان‌طور که در شکل زیر آمده است، لیبل هر نمودار را در ادامه آن نمایش دهیم. (۱ نمره)

۱. دو سری زمانی ایجاد کنید و لیبل‌های آن‌ها را (مانند شکل زیر) در پایان سری‌ها نمایش دهید.



۳. داده‌های مربوط به امضای آنلاین چند نفر در مسیر این data/signature قرار داده شده که شامل اطلاعاتی نظیر مختصات x و y، فشار، فاصله زمانی و جهت‌های زاویه‌ای قلم (متشکل از azimuth و altitude) است. (۱۰ نمره)

^۴ Matplotlib

^۵ Time Series



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

امیرمحمد کوبش پور

دکتر سامان هراتی زاده

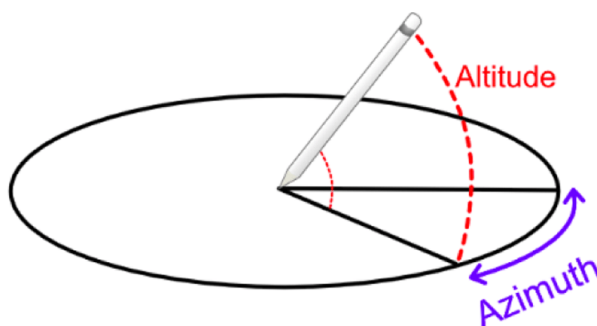
دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

index	x	Y	Time	azimuth	altitude	pressure
0	1,240	7,836	0	114	68	274
1	1,266	7,830	10	116	68	404
2	1,266	7,844	20	117	67	420

- این داده‌ها را با استفاده از کتابخانه‌های لازم، همانند جدول بالا (به ترتیب Header ها توجه کنید) در آورید و خروجی را در قالب فایل اکسل گزارش کنید.
 - با توجه به اطلاعات دیتاست سعی کنید امضای فرد را به صورت بصری در بیاورید. در این فرآیند ابتدا از دو ویژگی x و y برای ایجاد یک تصویر که در واقع تصویر امضای یک فرد است استفاده کنید. تصویر امضای خروجی حاصل این مرحله را در گزارش خود درج کنید.
 - با استفاده از ویژگی فشار (pressure) در یک بازه مشخص رنگی، میزان فشار وارده توسط کاربر را در تصویر خروجی نمایش دهید.
 - سپس بررسی کنید، اضافه کردن چه ویژگی‌هایی از این دیتاست مانند زاویه‌های قلم نسبت به صفحه، بر روی واقعی شدن تصویر خروجی تاثیرگذار است. (برای کسب اطلاعات بیشتر به بخش توضیحات این سوال مراجعه کنید، همچنین مطالعه بیشتر در مورد تاثیرگذاری این ویژگی‌ها به مرجع [۱] مراجعه کنید) تصاویر خروجی این مرحله را گزارش کنید و در مورد تاثیرگذاری این ویژگی‌ها بحث کنید.
 - مانند داده‌های این تمرین یک داده با توجه به تصویر امضای خود که باید شامل x, y باشد آماده کنید و آن را مصور کنید. در گزارش ارسالی، داده جمع آوری و تصویر امضای اصلی خود را ارسال کنید.
- توضیحات: در شکل زیر زاویه‌های azimuth و altitude قابل مشاهده است. زاویه altitude مربوط به زاویه قلم نسبت صفحه نمایشگر



شکل ۲: زاویه‌های azimuth و altitude که هنگام گرفتن قلم ذخیره شده‌اند [2].

است که در دادگان موجود در محدوده ۳۰ درجه تا ۹۰ درجه است، همچنین azimuth زاویه مربوط به محدوده حرکتی قلم است که مقدار آن می‌تواند از ۰ درجه تا ۳۶۰ درجه باشد.



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

امیرمحمد کویش پور

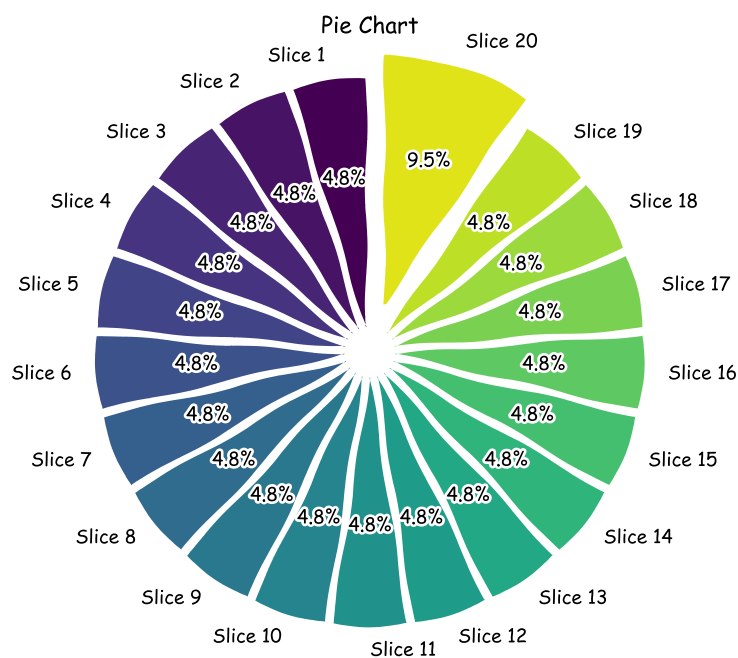
دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

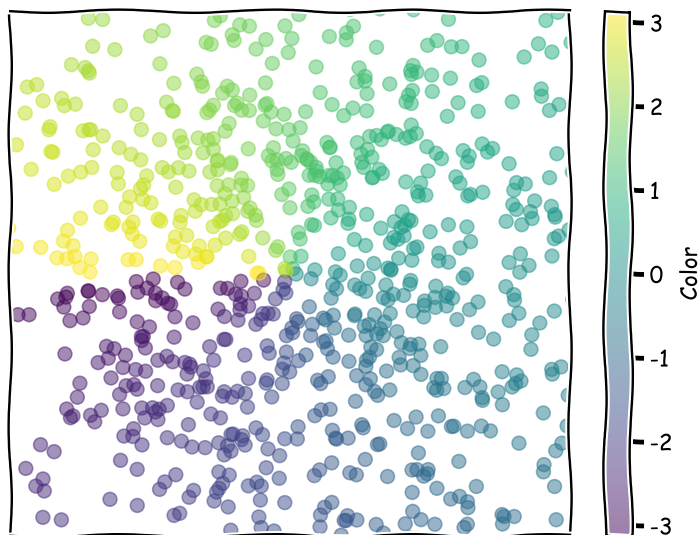
نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

۴. تصاویر زیر را به کمک matplotlib پیاده سازی و گزارش کنید. (۲۴ نمره)



شکل 3: نمودار دایره ای



شکل 4: نمودار پراکندگی



یادگیری ماشین

تمرین «صفر»

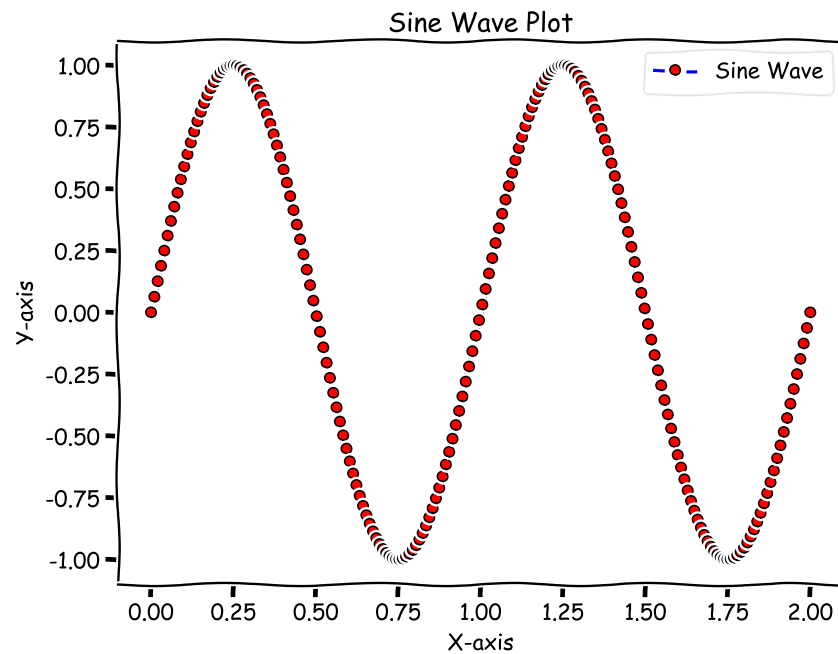
دستیاران آموزشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

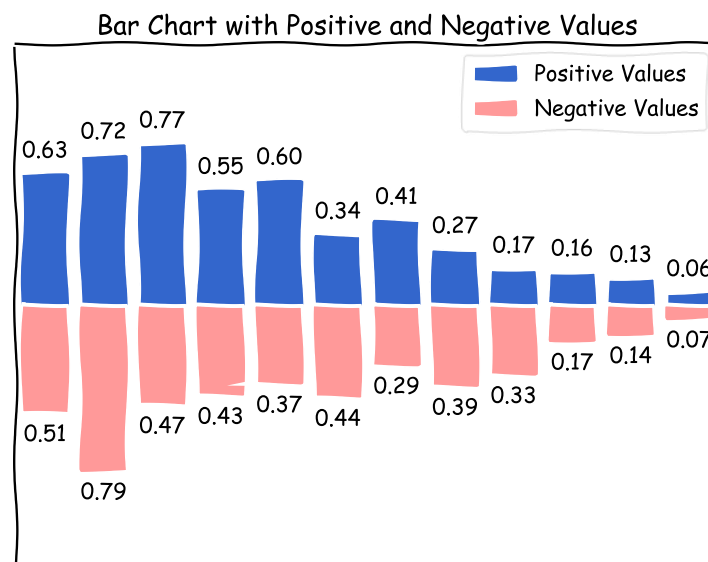
دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۲-۱۴۰۳

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲



شکل 5: نمودار موج سینوسی



شکل 6: نمودار میله ای



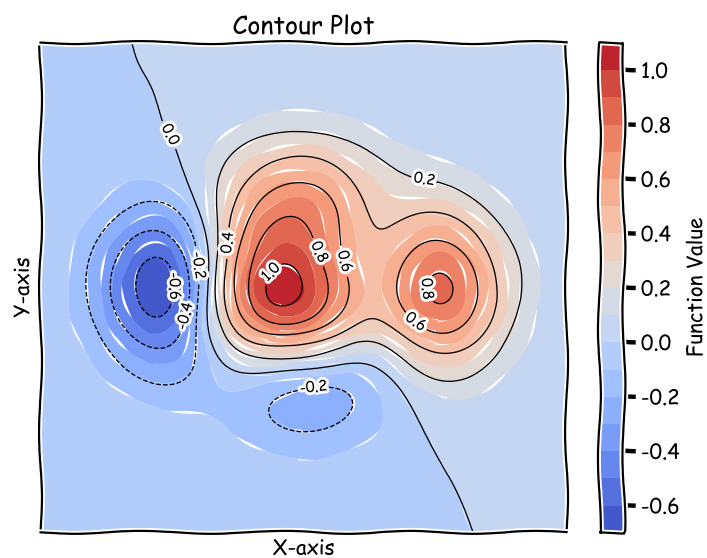
یادگیری ماشین

تمرین «صفر»

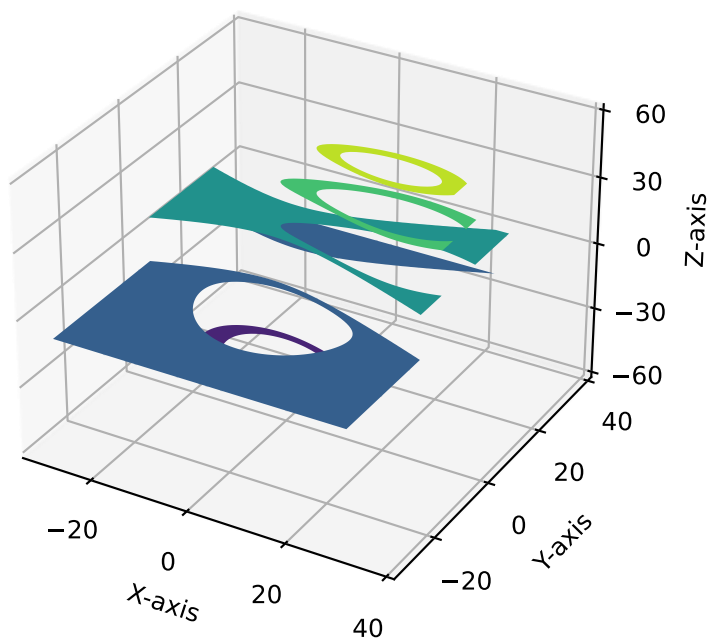
دستیاران آموزشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده
دانشگاه تهران - دانشکده سامانه های هوشمند
نیم سال اول ۱۴۰۲-۱۴۰۳

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲



شکل 7: کانتور



شکل 8: کانتور سه بعدی



یادگیری ماشین

تمرین «صفر»

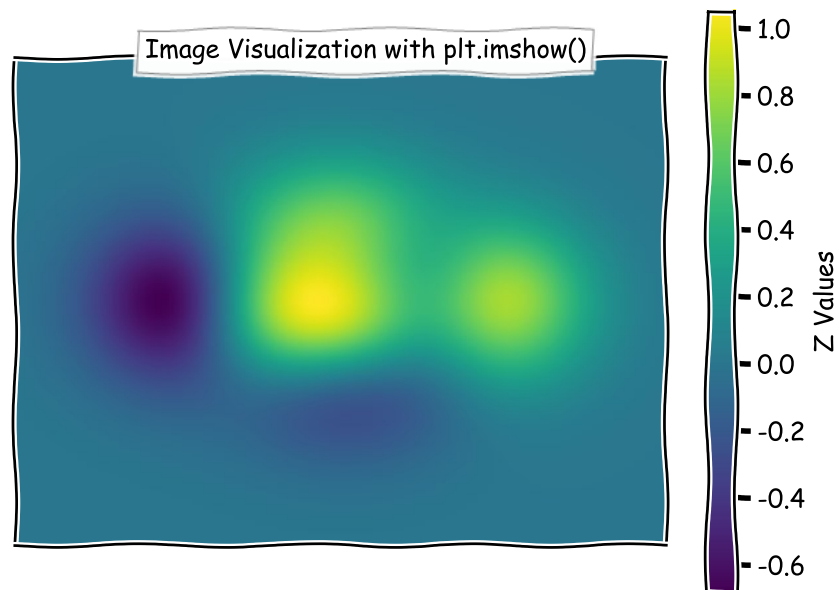
دستیاران آموزشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

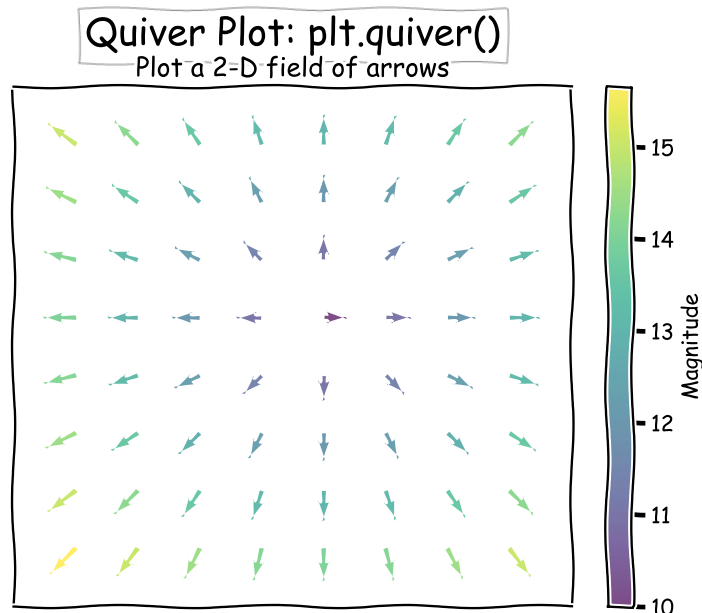
دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲



شکل 9: مصورسازی نگاره با imshow



شکل 10: نمودار Quiver



یادگیری ماشین

تمرین «صفر»

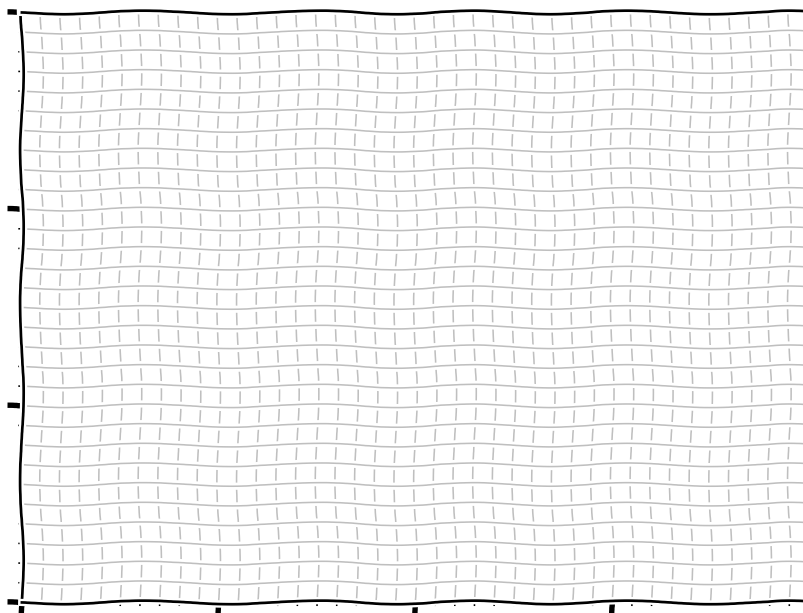
دستیاران آموزشی
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲



شکل ۱۱: صفحه شبکه ای



شکل ۱۲: زیرنمودارها به صورت گروهی



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

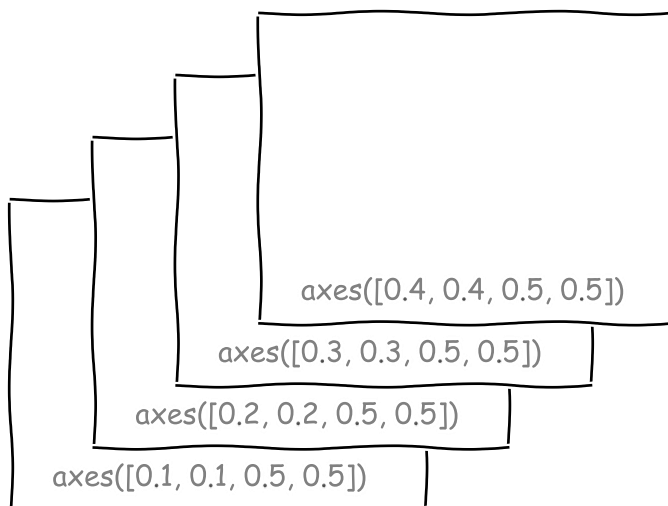
امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه های هوشمند

نیم سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲



شکل ۱۳: زیرنمودارها به صورت گروهی

شکل ۱۴: رسم متن

برای مشاهده نشانگرها^۶ و نگاشت رنگها^۷ به مستندات مت پلات لیب مراجعه کنید.

⁶ [Markers](#)

⁷ [Colormaps](#)



یادگیری ماشین

تمرین «صفر»

دستیاران آموزشی

امیرمحمد کوشش پور

دکتر سامان هراتی زاده

دانشگاه تهران - دانشکده سامانه‌های هوشمند

نیم‌سال اول ۱۴۰۳-۱۴۰۲

ساعت ۲۳:۵۹ | ۱۵ آبان ۱۴۰۲

سای پای و سیم پای

بخشی از دادگان حقوق قشر دانشگاهی در مسیر data/salary آورده شده است. با اجرای سلول مربوطه دیتافریم مرتبط بارگذاری می‌شود. قصد داریم با کمک آزمون‌های فرض آماری^۸ دست به قضاوت بزنیم! در این تمرین قصد داریم به کمک کتابخانه‌های مرتبط به مقایسه میانگین درآمد دو جامعه بپردازیم. در اینجا قصد داریم از $T\text{-test}$ کمک بگیریم.

۱. فرض می‌شود که اساتید خانم در دانشگاه از حقوق بالاتری برخوردار هستند، نتیجه‌ی بررسی محاسبات و تصمیم‌گیری خود را در گزارش ارائه کنید. (۱ نمره)

قصد داریم که کمی با کتابخانه‌های مناسب عملیات ریاضی آشنا شویم.

۲. از عبارت $f(x) = x^3 - 3x^2 + y^3 - 3y^2$ گرادیان و مشتق دوم بگیرید و نتیجه را گزارش کنید. (۱ نمره)

⁸ Statistical Hypothesis Testing