



تمرین اول

۱ مقدار جبر خطی

I ضرب و ابعاد

ماتریس‌های زیر را در نظر بگیرید.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix}$$

آ

توضیح دهید که چگونه می‌توان بدون ضرب کردن ماتریس‌ها، ابعاد ماتریس‌های زیر را به‌دست آوریم. ابعاد ماتریس‌ها را بنویسید.

$$\mathbf{BA}, \mathbf{B}^T, \mathbf{B}^T \mathbf{A}, \mathbf{A}^T \mathbf{B}$$

ب

ماتریس‌های بخش قبل را محاسبه کنید.

II ضرب و ابعاد، دشوارتر

آ

اگر $\mathbf{x}^{(i)} \in \mathbb{R}^{1 \times 2}$ یعنی $\mathbf{x}^{(i)} = \begin{bmatrix} x_1^{(i)} & x_2^{(i)} \end{bmatrix}$ باشد، و از طرفی $\vec{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$ ، ابعاد $\mathbf{x}^{(i)} \vec{\theta}$ چیست؟ برای نشان دادن تمامی $\mathbf{x}^{(i)}$ ‌ها از ماتریس $\mathbf{X} \in \mathbb{R}^{n \times 2}$ استفاده می‌کنیم (یعنی تعداد n نمونه از $\mathbf{x}^{(i)}$ داریم). ابعاد $\mathbf{X} \vec{\theta}$ چیست؟
نکته: نمادگذاری^۱ یک کمیت نرده‌ای^۲ به صورت یک حرف لاتین کوچک کج^۳ (a) ، یک بردار به صورت حرف لاتین کوچک^۴ پرننگ^۵ یا با یک پیکان^۶ روی آن (\vec{a}) یا (\mathbf{a}) و نمادگذاری یک ماتریس با حرف لاتین بزرگ^۷ پرننگ^۷ (\mathbf{A}) است.

^۱ notation

^۲ scalar

^۳ italic

^۴ lowercase letter یا small letter

^۵ bold font

^۶ arrow

^۷ uppercase letter یا capital letter

ب

می‌دانیم $\mathbf{z}^T \mathbf{z} = \sum_{i=1}^n (z^{(i)})^2$. یعنی می‌توان مجموع مربعات یک مجموعه از اعداد را به صورت ضرب بردار آن اعداد نوشت. حال مسئله زیر را در نظر بگیرید:

عبارت $J = \sum_{i=1}^n (\mathbf{x}^{(i)} \vec{\theta} - y^{(i)})^2$ را داریم. در مرحله اول J را به صورت ضرب دو بردار بازنویسی کرده و سپس گرادین آن را نسبت به بردار θ به دست آورید. $(\nabla_{\theta} J)$

۲ کمی آمار و احتمالات

I مقدمه

احتمالات بیزی^۱ را شرح دهید.

II سناریو

آ پرده اول

علی روز امتحان از خواب بیدار می‌شود و حال خوشی ندارد. او حتی توان رفتن به دانشگاه را هم ندارد و نمی‌تواند به امتحان ترم خود برسد. او ناچار به پزشک مراجعه می‌کند. پزشک برای علی یک آزمایش تجویز می‌کند. هفته بعد، هم‌زمان با اینکه نمره‌ها در سیستم ثبت شده و علی در درسش مردود شده است، نتیجه آزمایش نیز می‌رسد. آزمایش علی برای یک سرطان نادر مثبت شده است. بیماری‌ای که از هر ده‌هزار نفر در جهان، یک نفر به آن مبتلا می‌شود:

$$P(\text{sick}) = \frac{1}{10^4}$$

خبر بسیار نگران‌کننده است، اما پزشک به علی اعلام می‌کند که از هر صد نفری که آزمایششان مثبت می‌شود، ۹۹ نفر واقعاً دچار این بیماری هستند:

$$P(+|\text{sick}) = \frac{99}{100}$$

در میان آشفتگی‌ای که علی هم در امتحان مردود شده بود و هم آزمایش او با صحت^۲ ۹۹ درصد نتیجه سرطان را مثبت نشان داده بود، عمو محمد از راه می‌رسد و می‌گوید:

حالا که تست علی مثبت شده، احتمال اینکه علی این بیماری رو داشته باشه 0.980392 درصده
 $(P(\text{sick}|+) = 0.00980392)$.

به علی توضیح دهید که عمو محمد چگونه به چنین نتیجه‌ای رسید.

ب پرده دوم

عمو محمد آزمایشگاهی را سراغ دارد که همین آزمایش را با صحت 99.99 درصد انجام می‌دهد:

$$P(+|\text{sick}) = 0.9999$$

او علی را مجاب می‌کند تا دوباره آزمایش دهد. ترم جدید شروع می‌شود و علی به جای نشستن سرکلاس درس، روی صندلی آزمایشگاه نشسته و خون می‌دهد. روز حذف و اضافه، نتیجه آزمایش می‌رسد و از شانس بد علی، نتیجه دوباره مثبت است.

Bayesian probability^۱
 sensitivity^۲

در شرایطی که دو بار نتیجه آزمایش مثبت شده است، چقدر احتمال دارد که علی سرطان داشته باشد؟

$$P(\text{cancer} | ++)=?$$

ج پرده سوم

اوضاع عجیبی برای علی رقم خورده است. او مجبور به حذف پزشکی ترم خود می‌شود. علی پشت سیستم نشسته و به‌عنوان آخرین بازی ویدیویی زندگی خود، League of Legends را بازی می‌کند، که خود سرطانی دیگر است. ناگهان تلفن همراه او زنگ می‌خورد. مسئول آزمایشگاه دوم با علی تماس گرفته و به او می‌گوید که از آزمایش قبل خون اضافه مانده بود و بار دیگر آزمایش را انجام داده‌اند. او می‌گوید که در آزمایش جدید نتیجه منفی شده است.

از آنجایی که علی تاکنون آمار بیزی را به‌خوبی فرا گرفته است، از مسئول آزمایشگاه صحت آزمایش را می‌پرسد تا خود احتمال ابتلا به سرطان را محاسبه کند. اگر صحت آزمایش 99.9999 درصد باشد:

$$P(-|\text{cancer}) = 0.999999$$

علی احتمال سرطان داشتن خود را به شرط تمامی این آزمایش‌ها را چند درصد بدست می‌آورد؟

$$P(\text{cancer} | ++-)=?$$

۳ پردازش داده

CWRU Dataset I

در این بخش با دادگان^۱ عیب‌ارتعاشی یاتاقان^۲ Case Western Reserve University آشنا خواهید شد. سپس پردازش سیگنال را تمرین خواهید کرد. در این حین از کتابخانه‌هایی همچون matplotlib و Pandas استفاده می‌کنید.

آ دریافت داده

وارد این پیوند شده و یک فایل دلخواه از جدول موجود را دانلود کنید.

۱. فرمت فایل دانلودشده چیست؟ با چه دستوری می‌توان این فایل را در پایتون خواند؟
۲. ابتدا فایل را خوانده و در یک متغیر ذخیره کنید. توضیح دهید که نوع داده^۳ خوانده‌شده چیست؟ اجزای آن را نام ببرید.
۳. از میان سیگنال‌های موجود، یک سیگنال را انتخاب کرده و در متغیری دیگر ذخیره کنید.

ب نمایش سیگنال

۱. سیگنال را با استفاده از کتابخانه matplotlib نمایش دهید. محور افقی باید زمان را نمایش دهد. (فرکانس نمونه‌برداری 48KHz است.)
۲. سیگنال را برای بازه‌ی زمانی 2s تا 2.01s نمایش دهید.

ج تحلیل فرکانسی

۱. با استفاده از تبدیل فوریه، طیف فرکانسی سیگنال را نمایش دهید. (برای این کار، یک تابع بنویسید که سیگنال را در حوزه‌ی زمان به عنوان ورودی دریافت کرده و محاسبات را به صورت خودکار انجام دهد.)

^۱ dataset
^۲ bearing
^۳ data type

۲. فرکانس غالب^۱ چند هرتز است؟

د تقسیم‌بندی سیگنال

سیگنال را به قطعه‌هایی^۲ با اندازه ۱۲۸ تقسیم کرده و قطعه‌ها را در یک آرایه‌ی numpy ذخیره کنید. (هر سطر آرایه باید یک قطعه را شامل شود.) مقدار هم‌پوشانی^۳ قطعه‌ها دلخواه است.

ه ادامه تقسیم‌بندی سیگنال

۱. نمونه‌های^۴ به‌دست‌آمده را در یک pandas.DataFrame ذخیره کنید.
۲. با استفاده از DataFrame ایجادشده، ۱۰ قطعه‌ی مضرب^۵ ۱۳ را در یک نمودار روی هم نمایش دهید. نمودارها باید رنگ‌های مختلفی داشته و نمودار نهایی باید دارای برجسب^۶ باشد. (تمامی این مراحل باید در یک حلقه‌ی for انجام شود.)

و استخراج ویژگی (Feature Extraction)

در یادگیری ماشین^۷، برای وظایفی مانند طبقه‌بندی^۸ سیگنال، معمولاً از خود داده‌ی خام به علت ابعاد بالا استفاده نمی‌شود. به همین دلیل، ویژگی‌هایی از داده استخراج می‌شوند. در پردازش سیگنال، ویژگی‌های آماری می‌توانند مفید باشند.

۱. تابعی بنویسید که یک سیگنال را به عنوان ورودی دریافت کرده و مقادیر میانگین^۹، انحراف معیار^{۱۰}، و ریشه‌ی میانگین مربعات^{۱۱} را محاسبه کرده و بازگرداند.
۲. از تابع ایجادشده استفاده کرده و تمامی ویژگی‌ها را برای تمامی نمونه‌ها محاسبه کنید. مقادیر به‌دست‌آمده را در یک pandas.DataFrame جدید ذخیره کنید. (نام هر ویژگی را به عنوان نام ستون در DataFrame درج کنید.)
۳. DataFrame جدید را در یک فایل csv ذخیره کنید. دقت کنید که شماره نمونه‌ها نباید به عنوان یک ستون جدا باشد و نام ستون‌ها در فایل نهایی درج شود.

II Iris Dataset

در این بخش تمرکز بیشتری روی کتابخانه‌ی pandas خواهیم داشت و از دستورات جدیدی از کتابخانه‌ی matplotlib استفاده خواهیم کرد. همچنین از دادگان گل زنبق^{۱۲} استفاده می‌کنیم که یکی از معروف‌ترین دادگان در حوزه‌ی هوش مصنوعی است.

آ بررسی اولیه

۱. درباره‌ی این دادگان تحقیق کرده و بخش‌های مختلف آن را توضیح دهید.
۲. با استفاده از کتابخانه‌ی scikit-learn این دادگان را فراخوانی کنید.
۳. دادگان را به بخش‌های آموزش^{۱۳} و آزمون^{۱۴} تقسیم کنید.

dominant ^۱
segment ^۲
overlap ^۳
sample ^۴
label ^۵
machine learning ^۶
classification ^۷
mean ^۸
standard deviation ^۹
root mean squared (RMS) ^{۱۰}
Iris ^{۱۱}
trian ^{۱۲}
test ^{۱۳}

۴. داده‌ها را به یک `pandas.DataFrame` تبدیل کرده و نام هر یک از ستون‌ها را مشخص کنید. (این کار را هم برای دادگان آموزش انجام دهید و هم برای دادگان آزمون)
۵. سپس دو `DataFrame` آموزش و آزمون را با هم ادغام کنید. حتماً یک ستون در `DataFrame` نهایی وجود داشته باشد که داده‌های آموزش و آزمون را مشخص کند.

ب تحلیل بصری داده‌ها

۱. دو ویژگی دلخواه انتخاب کرده و پراکندگی نمونه‌ها را به صورت دوبعدی و با تفکیک نوع گل زنبق نمایش دهید.
۲. سه ویژگی دلخواه انتخاب کرده و پراکندگی نمونه‌ها را به صورت سه‌بعدی و با تفکیک نوع گل زنبق نمایش دهید.
۳. نقشه‌ی حرارتی^۱ ویژگی‌ها را نمایش دهید. (از کتابخانه‌ی `seaborn` برای نمایش نقشه حرارتی استفاده کنید).
۴. تابع چگالی احتمال^۲ ویژگی‌های دادگان را به تفکیک داده‌ی آموزش و آزمون نمایش دهید. (هر ویژگی باید در یک نمودار جداگانه نشان داده شود، و داده‌های آموزش و آزمون باید روی همان نمودار نمایش داده شوند).

ج گسسته‌سازی پیوستگی

یکی از ویژگی‌های عددی را انتخاب کرده و مقادیر عددی پیوسته را به چند کلاس گسسته تبدیل کنید. سپس آن را به عنوان یک ویژگی جداگانه در `DataFrame` ذخیره کنید. (مقادیر گسسته‌ی جدید می‌توانند دارای سه کلاس مانند "کوتاه"، "متوسط"، و "بلند" باشند).

د تحلیل آماری

ویژگی‌های آماری مختلف را برای کلاس `Setosa` توصیف کنید. (از متد `describe()` بهره بگیرید).

^۱ heatmap
^۲ probability density function

در انجام این تمرین حتماً به نکات زیر توجه کنید:

- موعد تحویل این تمرین، ساعت ۱۸:۰۰ روز پنج‌شنبه ۹ اسفندماه ۱۴۰۳ است.
- برای گزارش لازم است که پاسخ هر سوال و زیربخش‌هایش به‌ترتیب و به‌صورت مشخص نوشته شده باشند. بخش زیادی از نمره به توضیحات دقیق و تحلیل‌های کافی شما روی نتایج بستگی خواهد داشت.
- لازم است که در صفحه اول گزارش خود لینک مخزن گیت‌هاب را و گوگل‌کولب مربوط به مینی‌پروژه خود را درج کنید. درخصوص گیت‌هاب، یک مخزن خصوصی درست کنید و آی‌دی‌های MJAHMADEE و AliBagheriNejad را به‌عنوان Collaborator به مخزن اضافه کنید. پروژه‌های گیت‌هاب می‌بایست در انتهای ترم پابلیک شوند. درمقابل، لینک گوگل‌کولب را در حالتی که دسترسی عمومی دارد به اشتراک بگذارید. دفترچه‌کد گوگل‌کولب باید به‌صورت منظم و با بخش‌بندی مشخص تنظیم شده باشد و خروجی سلول‌های اجراشده قابل مشاهده باشد. در گیت‌هاب نیز یک مخزن برای درس و یک پوشه مجزا برای هر مینی‌پروژه ایجاد کنید.
- (آموزش پرایوت‌کردن مخزن گیت‌هاب و آموزش افزودن Collaborator به مخزن گیت‌هاب)
- هر جا از دفترچه‌کد گوگل‌کولب شما نیاز به فراخوانی فایلی خارج از محیط داشت، مطابق آموزش‌های ارائه‌شده ملزم هستید از دستور `gdown` استفاده کنید و مسیرهای فایل‌ها را طوری تنظیم کنید که صرفاً با اجرای سلول‌های کد، امکان فراخوانی و خواندن فایل‌ها توسط هر کاربری وجود داشته باشد.
- در تمامی مراحل تعریف داده و مدل و هر جای دیگری که مطابق آموزش‌های ویدیویی و به لحاظ منطقی نیاز است، Random State را برابر با دو رقم آخر شماره دانشجویی خود در نظر بگیرید.
- استفاده از ابزارهای هوشمند (مانند ChatGPT) در کمک‌گرفتن برای بهبود کدها مجاز است؛ اما لازم است تمام جزئیات مواردی که در خروجی‌های مختلف گزارش خود عنوان می‌کنید را به خوبی خوانده، درک و تحلیل کرده باشید. استفاده از این ابزارهای هوشمند در نوشتن گزارش و تحلیل‌ها ممنوع است.
- در جاهایی که با توجه به دو رقم آخر شماره دانشجویی خود محدود به انتخاب عدد، متغیر و یا داده‌ای خاص شده‌اید، برای تست‌های اضافه‌تر و نمایش بهبود در نتایج خود، مجاز هستید از مقادیر دیگر هم استفاده کنید.
- رعایت نکات بالا به حرفه‌ای‌تر شدن شما کمک خواهد کرد و اهمیتی معادل مطالب درسی فراگرفته‌شده دارد؛ بنابراین، در صورت عدم رعایت هریک از این نکات، از نمره تمرین شما کسر خواهد شد.
- آی‌دی پرسش هرگونه سوال درخصوص مینی‌پروژه شماره 0.1