

بنام خدا

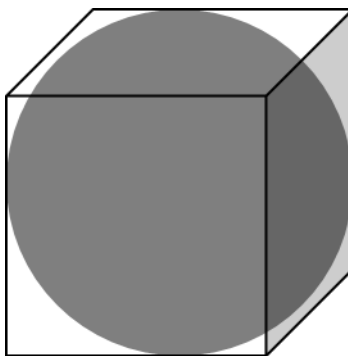
دانشگاه یزد - بهار ۱۴۰۳

تمرین کامپیوتری شماره ۱ درس بازشناسی الگوهای آماری

پاسخ سوالات ذیل را در فایل notebook پیوست بنویسید. دقت کنید که پاسخ هر سوال در قسمت مرتبط با همان سوال نوشته شود.

به توضیحات نوشته شده در هر قسمت توجه نمایید.

۱. برنامه‌ای بنویسید که شعاع کره شکل زیر را به عنوان ورودی بگیرد و حجم ناحیه خارج از کره و داخل مکعب را تا ۶ رقم اعشار محاسبه و به عنوان خروجی اعلام کند.



۲. برنامه‌ای بنویسید که از کاربر بخواهد تا یک عدد صحیح مثبت را که زمان بر حسب روز است وارد کند و سپس برنامه، زمان وارد شده را به سال، ماه، هفته، روز و ساعت تبدیل و اعلام کند.

۳. دو ماتریس A و B را به صورت زیر در نظر بگیرید و عملیات‌های زیر را یک بار با استفاده از کتابخانه Numpy و بار دیگر بدون استفاده از تابع آماده آن عملیات در Numpy، انجام دهید.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 7 & 6 & 4 \end{bmatrix}$$

$$B = \begin{bmatrix} 5 & 2 & 6 \\ 5 & 6 & 7 \\ 7 & 6 & 4 \end{bmatrix}$$

الف) ضرب ماتریسی بین دو ماتریس A و B را یک بار بدون استفاده و یک بار با استفاده از حلقه for محاسبه کنید.
ب) دترمینان ماتریس A را به دست آورید.

۴. دو ماتریس قسمت قبل را در نظر بگیرید و با استفاده از توابع آماده Numpy:

الف) ضرب خارجی دو ماتریس را محاسبه کنید
ب) ضرب داخلی دو ماتریس را محاسبه کنید.

ج) معکوس دو ماتریس A و B را به دست آورید.
د) مقادیر Eigen value را برای دو ماتریس A و B به دست آورید.

۵. با استفاده از کتابخانه‌های Numpy و Matplotlib یک تابع سیگموئید را برای مقادیر بین ۱۰- تا ۱۰ با فاصله ۰.۲ رسم نمایید.

۶. ابتدا با استفاده از بخش random کتابخانه Numpy مجموعه‌ی ۱۰۰۰۰ نمونه‌ای از اعداد تصادفی با توزیع گاوسی با میانگین ۱۰ و واریانس ۱ تولید نمایید. سپس با توجه به آن برنامه‌های زیر را بنویسید:
الف) هیستوگرام اعداد تولید شده در بخش پیشین را توسط کتابخانه Matplotlib رسم نمایید.
ب) با استفاده از دستور sample در کتابخانه Numpy دویست نمونه تصادفی از بردار اول را انتخاب نمایید.
ج) هیستوگرام اعداد نمونه‌برداری شده را رسم نموده و در مورد تفاوت آن با هیستوگرام بخش «ب» بحث کنید.
د) یک مجموعه ۱۰۰۰۰ نمونه‌ای از اعداد تصادفی با توزیع یکنواخت بین ۱- و ۱ تولید کرده و هیستوگرام آن را رسم کنید.
ه) مجموعه اعداد بردار اول و قسمت «د» را در هم ضرب کرده و هیستوگرام حاصل را رسم کنید. در مورد توزیع این دادگان بحث کنید.

۷. ابتدا با استفاده از کتابخانه Pandas فایل books.csv را خوانده و در یک dataframe ذخیره کنید. سپس با توجه به به داده‌های مذکور، موارد خواسته شده را پیاده‌سازی کنید.
الف) ۵ داده اول را نمایش دهید.

ب) ستون‌های غیرضروری را حذف کرده و تنها ستون‌های زیر را حفظ کنید

isbn, original_publication_year, original_title, authors, ratings_1, ratings_2, ratings_3,
ratings_4, ratings_5

ج) نام ستون‌ها ذکر شده را به عناوین زیر تغییر دهید.

ISBN, Publication Year, Original Title, Authors, One Star Reviews, Two Star
Reviews, Three Star Reviews, Four Star Reviews, Five Star Reviews

د) یک نمای کلی ساده و تحلیلی از ستون‌های عددی مجموعه داده ارائه دهید.

ه) dataframe جدید را در یک فایل csv. دیگر با عنوان "book_clean.csv" ذخیره کنید و سوالات بعدی را با استفاده از dataframe جدید پیاده‌سازی کنید.

و) اسامی نویسندگان در ستون "author" را چاپ کنید.

ز) مشخص کنید هر نویسنده چند کتاب به چاپ رسانده است.

ح) کتابی که زودتر از سایر کتاب‌ها به چاپ رسیده است را مشخص کنید.

ط) کتابی که دیرتر از سایر کتاب‌ها به چاپ رسیده است را مشخص کنید.

ی) تعداد کل نظرات موجود در dataframe را به دست آورید.

۸. یک مسئله طبقه‌بندی دو کلاسه با بردار ویژگی دوبعدی $X = (x_1, x_2)^t$ را در نظر بگیرید. ω_1 و ω_2 دو کلاس مذکور هستند و

$$p(X|\omega_1) \sim N\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \Sigma_1\right),$$

$$p(X|\omega_2) \sim N\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \Sigma_2\right),$$

$$p(\omega_1) = p(\omega_2) = \frac{1}{2}$$

به صورتی که:

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$$

با توجه به داده‌های بالا مسائل زیر را پاسخ دهید:

الف) مرز تصمیم بیز را محاسبه کنید.

ب) به صورت تصادفی، ۵۰ الگو از هر یک از چگالی‌های شرطی دو کلاسه (two class-conditional density) را در یک فضای ویژگی دوبعدی رسم کنید. همچنین مرز تصمیم که در قسمت «الف» به دست آورده‌اید را هم ترسیم کنید.

ج) هزار الگوی تست از هر کلاس تولید کرده و نرخ خطای empirical را بر اساس مرز به دست آمده در قسمت «الف» محاسبه کنید.

۹. دوست گیاه‌شناس شما دریافته است که شما درس یادگیری را اخذ نموده‌اید و از این رو از شما کمک می‌خواهد. او دو نوع

قارچ در جنگل یافته است که یکی از آن‌ها به شدت سمی است (Agaric Moribundus) و دیگری (Agaric Stimulantus) هم شامل سم و هم شامل دارویی مهم برای درمان یک بیماری مهم است. تفاوت این دو قارچ در ترکیبات شیمیایی آن‌ها است. دوست شما یکسری اندازه‌گیری غیرمستقیم و نویزی از مقادیر مواد شیمیایی این دو قارچ یعنی سم x و ماده فعال y را در اختیار شما قرار می‌دهد.

بر اساس این اندازه‌گیری‌ها شما می‌توانید مقادیر مدل‌های دو کلاسه $p(x, y|1)$ و $p(x, y|2)$ که هر دو گاوسی با مقادیر زیر هستند را محاسبه کنید.

$$\mu_1 = (8,2), \mu_2 = (2,8)$$

$$C_1 = C_2 = \begin{bmatrix} 4.3 & 0 \\ 0 & 3.1 \end{bmatrix}, p(\omega_1) = p(\omega_2)$$

الف) مرزهای تصمیم بیزین را محاسبه کرده و آنها را رسم نمایید.

ب) احتمالات پسین (posterior probabilities) را محاسبه و رسم نمایید.