

به نام خدا



دانشکده مهندسی کامپیوتر

مبانی هوش محاسباتی

تمرین سوم

PCA – SVD – t sne – kmeans

دکتر مائده سادات طاهایی

زمستان 1403

دستیاران تمرین: آیسامیاهی نیا – سپهر رضوی – فاطمه باقری

- در صورت وجود هرگونه ابهام به طراح پیام دهید.
- انجام تمرین ها تک نفره می باشد.
- زبان برنامه نویسی پایتون است.
- موارد ارسال شده به صورت آنلاین تحویل گرفته خواهند شد.
- کل فایل محتوای ارسالی را داخل فایل زیپ قرار داده و نام آن را شماره دانشجویی خود قرار دهید.
- تاریخ ریلیز تمرین: 20 اردیبهشت ماه
- تاریخ تحویل تمرین: 2 خرداد ماه
- آیدی طراحان در تلگرام: @Aysa_ma @FatemehBagherii @s_sepehr_razavi
- تاخیر مجاز: 15 روز | تاخیر مجاز در هر تمرین: 6 روز
- **در این تمرین به طور استثنا، امکان استفاده از تاخیر وجود ندارد.**

- نمره کل این تمرین از 102 بوده و 2 نمره امتیازی است.

1. سوالات تئوری

سوالات خانم میاهی نیا @Aysa_ma

1. در الگوریتم k-means با استفاده از فاصله اقلیدسی¹، داده‌های زیر را به ۳ خوشه² تقسیم کنید:
- $A_8=(4,9)$, $A_7=(1,2)$, $A_6=(6,4)$, $A_5=(7,5)$, $A_4=(5,8)$, $A_3=(8,4)$, $A_2=(2,5)$, $A_1=(2,10)$
- فرض کنید نقاط اولیه³ یا مراکز اولیه خوشه‌ها⁴ به ترتیب A_1 ، A_4 و A_7 باشند. الگوریتم k-means را تنها برای ۱ دوره⁵ اجرا کنید. در پایان این دوره، موارد زیر را ارائه دهید: (12نمره)
- خوشه‌های جدید (یعنی داده‌هایی که به هر خوشه تعلق دارند)
 - مراکز جدید خوشه‌ها
 - یک فضای ۱۰ در ۱۰ ترسیم کنید که تمام ۸ نقطه در آن مشخص باشند، و پس از اجرای یک دوره، خوشه‌ها و مراکز جدید را نشان دهید.
 - چند دوره دیگر برای همگرایی نیاز است؟ نتیجه را برای هر دوره رسم کنید.
2. سه نقطه داده در فضای دوبعدی داده شده‌اند: (8نمره)
- $(1,1)$ ، $(2,2)$ و $(3,3)$
- مؤلفه اصلی اول چیست؟
 - اگر بخواهیم داده‌ها را با استفاده از مؤلفه اصلی به فضای یک‌بعدی project کنیم، واریانس داده‌های project شده چقدر است؟
 - برای داده‌های project شده در قسمت قبل، اگر آن‌ها را دوباره در فضای دوبعدی بازسازی کنیم، خطای بازسازی⁷ چقدر خواهد بود؟

¹ Euclidean distance

² cluster

³ seeds

⁴ centers of each cluster

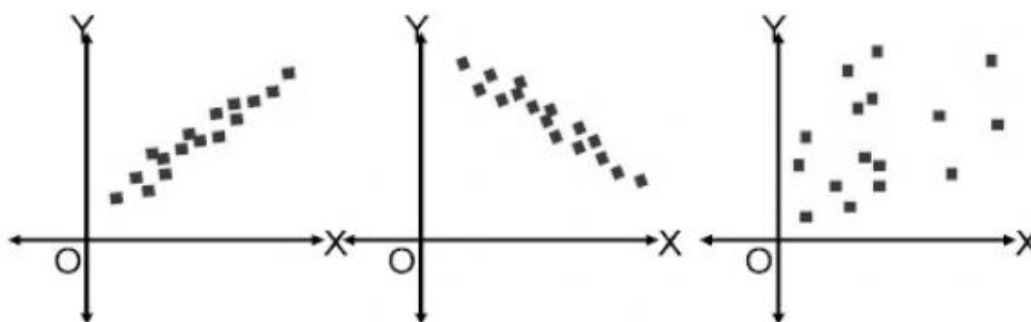
⁵ epoch

⁶ first principal component

⁷ reconstruction error

3. به سوالات زیر پاسخ دهید. (15 نمره)

- در چه شرایطی می توان برای انجام PCA، از هر دو روش SVD و covariance matrix استفاده کرد و نتیجه یکسان گرفت؟
- فرض کنید مجموعه داده ای دارید که شامل چند ویژگی مختلف است. برای بررسی رابطه بین برخی از این ویژگی ها، نمودار پراکندگی برای سه جفت ویژگی به صورت زیر رسم شده است:



با توجه به نمودارهای بالا، کدام یک از این نمودارها بیان گر وجود چندهم خطی⁸ بین ویژگی ها هستند؟ توضیح دهید چرا.

- در مسئله های کاهش ابعاد، دو تکنیک پرکاربرد یعنی PCA و t-SNE معمولاً استفاده می شوند. با توجه به تفاوت در رویکردهای ریاضی و نوع ساختاری که هر یک حفظ می کنند: تفاوت اصلی بین PCA و t-SNE در چیست؟ چه زمانی استفاده از هر کدام مناسب تر است؟

4. به سوالات زیر درباره الگوریتم K-means پاسخ دهید. (9 نمره)

- چرا در الگوریتم K-means فاصله اقلیدسی به فاصله منهتن⁹ ترجیح داده می شود؟
- آیا در الگوریتم K-means نیاز به Feature Scaling وجود دارد؟ با مثال توضیح دهید.
- برخی از موقعیت هایی را توضیح دهید که الگوریتم K-means در آن ها عملکرد خوبی ندارد.
- آیا مقداردهی اولیه به مراکز خوشه¹⁰ روی الگوریتم K-means تأثیر دارد؟

⁸ multicollinearity

⁹ Manhattan distance

¹⁰ centroid initialization

سوالات آقای رضوی @s_sepehr_razavi

5. SVD (12نمره)

۱. ثابت کنید ماتریس $A = 0$ باشد اگر و تنها اگر تمامی مقادیر ویژه هایش برابر صفر باشد.

۲. نشان دهید اگر ماتریس A ماتریسی وارون پذیر با مقادیر تکین s_1, s_2, \dots, s_n باشد، مقادیر تکین A^{-1} برابر با $s_1^{-1}, s_2^{-1}, \dots, s_n^{-1}$ خواهند بود.

۳. مقادیر ویژه ماتریس $A^T A$ مربع مقادیر تکین ماتریس A می باشند.

تجزیه SVD ماتریس زیر را پیدا کنید.

$$A = \begin{pmatrix} 3 & 3 & 2 \\ 2 & 3 & -2 \end{pmatrix} \quad (5)$$

سوالات خانم باقری @FatemehBagherii

6. در الگوریتم تحلیل مؤلفه های اصلی (PCA)، چرا انتخاب بردار ویژه متناظر با بزرگترین مقدار ویژه، بهترین جهت برای کاهش ابعاد داده محسوب میشود؟ این موضوع را از منظر کمینه سازی خطای بازسازی توضیح دهید.

(5نمره)



2. سوالات عملی

سوالات خانم میاهی نیا @Aysa_ma

1. در فایل نوت بوک **k-means.ipynb** (18نمره)

a. بخش هایی که با عنوان **"START CODE HERE"** یا مشابه آن مشخص شده اند را کامل کنید.

b. به تمام پرسش های مطرح شده در نوت بوک به صورت کامل، دقیق و همراه با تحلیل مناسب پاسخ دهید.

توجه داشته باشید که پاسخ ها و تحلیل های شما باید در فایل گزارش نهایی نیز آورده شود.

2. نوت بوک پیوست شده **PCA.ipynb** شامل مراحل اولیه PCA روی مجموعه داده Fashion

MNIST است. بخش های مشخص شده را کامل کرده و نمودارهای مورد نیاز را تحلیل کنید. (5نمره)

سوالات آقای رضوی @s_sepehr_razavi

3. نوت بوک t sne را تکمیل کنید. (18نمره)

آنچه تحویل داده میشود:

1. کد اجرایی تمرینات عملی
 2. پاسخ های تمرین تئوری در یک فایل PDF
- 📌 نکته بسیار مهم: 2 مورد بالا را zip کرده و نام آن را شماره دانشجویی خود گذاشته و فقط در کوئرا ارسال کنید.