



تکلیف شماره ۵

- قبل از شروع تمرین، فایل مربوط به قوانین حل و تحویل تمرین‌ها را مطالعه کنید.
- سوالات و مشکلات خود را درباره این تمرین می‌توانید در گروه تلگرامی درس یا با طراحان این تمرین مطرح کنید.
- پاسخ سوالات را در یک فایل پی‌دی‌اف^۱ قرار دهید.
- مهلت ارسال پاسخ‌ها: جمعه ۲۸ دی ساعت ۲۳:۵۹
- طراح این تمرین: [محمد فلاح](#)

۱۰۰ نمره

سوالات تئوری

سؤال ۱ (۳۰ نمره)

با در نظر گرفتن الگوریتم‌های خوشه‌بندی به سوالات زیر پاسخ بدهید.

(الف) الگوریتم‌های خوشه‌بندی فرضیات خاصی درباره داده‌ها و نحوه توزیع آن‌ها دارند. این فرضیات چگونه می‌توانند نتیجه خوشه‌بندی را تحت تاثیر قرار دهند؟ (دو الگوریتم DBSCAN و K-Means و روش محاسبه و هایپر پارامترهای هر یک را در نظر بگیرید).

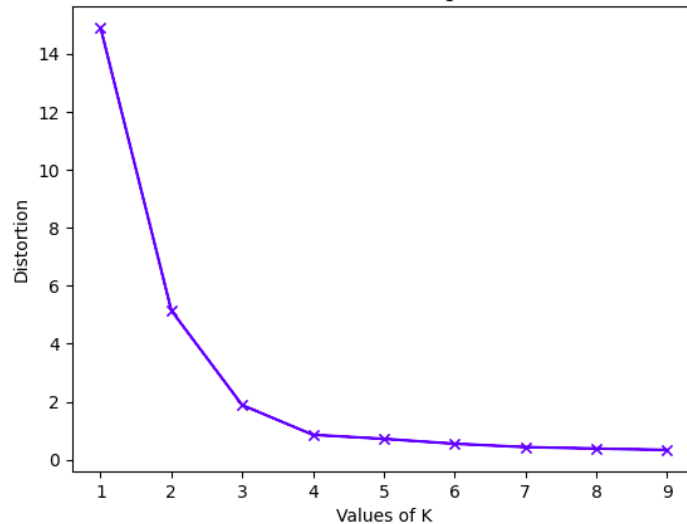
(ب) فرض کنید سه دیتاست با داده‌های دوبعدی داریم که در زیر نمایش داده شده‌اند. کدام الگوریتم در هر مورد خوشه‌بندی بهتری انجام می‌دهد؟



(ج) زیاد شدن ابعاد داده‌ها چگونه بر خوشه‌بندی تاثیر می‌گذارد؟

د) تصور کنید از الگوریتم DBSCAN استفاده می‌کنید و نتایج شما به تغییرات کوچک در پارامتر اپسیلون (ϵ) بسیار حساس هستند. این موضوع چه چیزی را درباره داده‌های شما نشان می‌دهد؟

ه) نمودار زیر میانگین مربع فاصله بین هر نقطه داده و مرکز خوشه تخصیص داده شده به ازای هر مقدار K در خوشه‌بندی K -Means را نشان می‌دهد. به نظر شما بهترین مقدار K برای این داده‌ها چیست؟



سؤال ۲ (۱۰ نمره)

به سوالات زیر درباره Support Vector Machine پاسخ دهید.

الف) داده‌های «غیرقابل جداسازی» به چه معناست؟ این موضوع چگونه چالش‌هایی برای روش پایه‌ای SVM ایجاد می‌کند؟

ب) چرا بیشینه کردن margin در SVM مهم است؟ margin چه چیزی را نشان می‌دهد؟

سؤال ۳ (۱۵ نمره)

فرض کنید یک متغیر تصادفی X از توزیعی با تابع چگالی احتمال (PDF) زیر پیروی می‌کند.

$$f(x; \alpha) = \{\alpha x^{\alpha-1} \text{ for } 0 < x < 1, 0 \text{ otherwise}\}$$

توجه به n نمونه مستقل $x_1, x_2, x_3, \dots, x_n$ برآورد maximum log likelihood برای پارامتر آلفا را محاسبه کنید.

سؤال ۴ (۱۳ نمره)

تجزیه SVD ماتریس زیر را محاسبه کنید.

3	2	2
2	3	-2

سؤال ۵ (۱۲ نمره)

هر قسمت تجزیه SVD یک ماتریس را توصیف کنید. به طور خاص، برای هر یک از ماتریس‌های U، S و V، موارد زیر را توضیح دهید:

الف) ابعاد آن‌ها در رابطه با ماتریس داده اصلی A (که ابعادی برابر با $m \times n$ دارد) چیست؟

ب) شهودی از معنای هر یک از ستون‌های آن‌ها می‌توان داشت؟

سؤال ۶ (۱۲ نمره)

به شما یک مجموعه داده از قدهای مردان و زنان داده شده است:

مردان: [۱۷۱, ۱۷۸, ۱۶۸, ۱۸۲, ۱۷۵]

زنان: [۱۶۳, ۱۷۰, ۱۵۸, ۱۶۵, ۱۶۲]

الف) با استفاده از MLE: توزیع‌های گاوسی جداگانه‌ای برای هر جنسیت پیدا کنید.

ب) احتمال اینکه یک فرد با قد ۱۶۹ سانتی متر مرد باشد را محاسبه کنید.

سؤال ۷ (۸ نمره)

چگونه احتمال $p(c|x)$ را در مدل بیز ساده محاسبه می‌کنیم و فرمول ساده شده آن چیست؟

سالم و موفق باشید.