



دانشکده مهندسی برق و کامپیوتر

## درس یادگیری ماشین

### پروژه کامپیوتری

نیمسال دوم سال تحصیلی ۱۴۰۰ - ۱۴۰۱



## سوالات

داده های بیان ژن نمونه های 5 سرطان BRCA, KIRC, COAD, LUAD and PRAD موجود است. این داده شامل 20531 مقدار بیان ژن (ستون) و 801 فرد (سطر) مختلف می باشد.

1) الگوریتم  $k\_means$  را با  $k=5$  و مقادیر مختلف اولیه اجرا کنید و بهترین نتیجه را گزارش دهید. همچنین مقادیر  $loss$ ,  $NMI$ ,  $ARI$  و  $purity$  را گزارش دهید. (۵ نمره)

2) با استفاده از الگوریتم  $k\_means$  داده ها را خوشه بندی نموده و مقدار بهینه  $k$  بنابر معیار سیلوئت گزارش کنید. (۵ نمره)

3) با استفاده از GMM داده ها خوشه بندی نموده و براساس معیارهای  $ARI$  و  $NMI$  و  $AMI$  با روش قبلی مقایسه شود. (۵ نمره)

4) در هر مورد از دو سوال بالا نتایج خوشه بندی به همراه خود داده ها توسط PCA در دو بعد نمایش دهید. (۵ نمره)



## نکات پیاده سازی و تحویل.

- انجام این پروژه به صورت یک نفره می باشد.
- گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است. لطفاً تمامی نکات و فرضیهایی که برای پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید را در گزارش ذکر کنید.
- از نوشتار محاوره‌ای و درج تصاویر دست نوشته در گزارش خودداری نمائید چون باعث کسر نمره خواهد شد.
- برای انجام پروژه استفاده از زبان برنامه‌نویسی پایتون الزامی می‌باشد.
- حتماً پیاده سازی خود را در محیط Jupyter Notebook و در یک فایل ipynb انجام دهید.
- نیازی به یک فایل پی دی اف جداگانه برای گزارش بخش پیاده‌سازی نیست. توضیحات خود را در همان فایل ipynb بنویسید. توضیحات می‌تواند به زبان فارسی و یا انگلیسی نوشته شوند. در صورت تحویل فایل جداگانه‌ای برای گزارش نمره‌ی این بخش اعمال نخواهد شد.
- در فایل Jupyter Notebook هر سوال از پروژه به همراه پاسخ آن مشخص شده و خروجی‌های مورد نیاز نیز ذخیره شده باشد. همچنین هرگونه نتیجه و یا تحلیلی که در شرح سوال از شما خواسته شده است را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- بعد از تکمیل پاسخ‌ها در فایل Jupyter Notebook مجدداً kernel را راه‌اندازی کرده و فایل را اجرا نمائید به صورتیکه شماره‌ی هر سلول در فایل دقیقاً مطابق با ترتیب سلول‌ها باشد.
- در صورت مشاهده تقلب امتیاز تمامی افراد شرکت کننده در آن، صفر لحاظ می‌شود.
- در صورت وجود سوال و یا ابهام می‌توانید از طریق آیدی تلگرام زیر با دستیار آموزشی در ارتباط باشید:

@hamidravaee

@basir\_ebr

@givkashi