



هدف پروژه:

یکی از مهم ترین علل ایجاد بیماری ها، ژن ها هستند. بیان یک ژن مخرب یا بیان نشدن یک ژن ضروری و یا حتی کم یا زیاد شدن بیان ژن هایی که در حالت عادی مخرب نیستند، میتوانند منجر به بیماری شوند. میزان بیان ژن های مختلف در بافت های بدن را میتوان از طریق روش های آزمایشگاهی اندازه گیری کرد. Microarray یک وسیله ی آزمایشگاهی است که با استفاده از آن میتوان میزان بیان ژن های مورد نظر را بررسی کرد. میتوان با تحلیل بر روی داده های به دست آمده از microarray و به دست آوردن تفاوت ها بین میزان بیان ژن در افراد بیمار و سالم، منشا احتمالی بیماری مورد نظر را مورد مطالعه قرار داد. در این پروژه قصد داریم با تحلیل روی داده های microarray متعلق به افراد مبتلا به سرطان خون و مقایسه ی آن با داده های افراد سالم، منشا این بیماری را مورد مطالعه قرار دهیم.

داده های مورد استفاده :

در این [لینک](#)، داده هایی که قرار است با آن کار کنید قرار دارند. این داده ها مربوط به نوعی از سرطان خون به نام AML یا سرطان خون حاد مغز استخوان هستند.

نیازمندی های لازم برای پیشبرد پروژه :

۱. از آنجا که این پروژه قرار است با زبان R انجام شود، در ابتدا لازم است زبان R و IDE مربوط به آن یعنی RStudio را نصب کنید.
۲. برای آموزش کار با داده ها و انجام تحلیل های مورد نیاز روی آن ها لازم است ویدیو های جلسات ۹ تا ۱۲ این [لینک](#) را مشاهده بفرمایید.

موارد قابل تحویل :

برای این پروژه باید موارد زیر تحویل داده شود :

۱. کد های زده شده
۲. گزارشی از روند کار و موارد خواسته شده (گزارش باید شامل نام و شماره دانشجویی تمام افراد گروه باشد و تمام فاز ها باید با گروه ثابت انجام گیرد)
۳. ددلاین این فاز تا پایان یکشنبه ۱۳ آذر خواهد بود و پس از آن ۳ روز تاخیر مجاز وجود دارد که از تعداد روزهای تاخیر از همه ی اعضای گروه کسر میشود.

فاز اول :

۱. درمورد microarray ، روش کار آن و فرمت داده های خروجی آن به طور مختصر توضیح دهید. (۱۰)

۲. داده هایی که phenotype آنها normal است را به عنوان داده های گروه سالم و داده هایی که source name آن ها AML patient است را به عنوان گروه داده های بیمار در نظر بگیرید. داده های اولیه ممکن است برای تحلیل های بعدی آماده نباشند. کیفیت داده ها را از جنبه هایی که به نظرتان میرسد بررسی کنید و در صورت لزوم تغییرات لازم را روی آن ها اعمال کنید. (راهنمایی: برای مثال نرمال سازی داده ها). برای هر ویژگی ای که کیفیتش را کنترل میکنید ذکر کنید که این کنترل چه لزومی دارد، مراحل بررسی و کنترل خود را گزارش کنید (برای مثال نمودارها و...) و اگر لازم بود تغییری در داده ها ایجاد کنیم، تاثیر تغییرات را گزارش کنید. (۳۰)

۳. لزوم کاهش ابعاد داده ها چیست؟ سه روش مختلف برای کاهش ابعاد را انتخاب کرده و نتایج حاصل از هر سه روش را گزارش کنید. سپس با مقایسه این نتایج، روشی که بهترین خروجی را نتیجه داده است، انتخاب کنید. دلایل انتخاب روش بهتر را ذکر کنید. راهنمایی: برای مثال می توانید کاهش ابعاد را با سه روش PCA، MDS و tSNE انجام دهید. (۳۰)

۴. اگر دقت کنید source name نمونه های نرمال مختلف با یکدیگر تفاوت دارند. این فیلد بیانگر چیست؟ اگر داده ها را بر اساس source name گروه بندی کنیم (همه ی نمونه های بیمار در یک گروه هستند و هر نمونه ی سالم در گروه source name متناظر با خود است) همبستگی بین گروه ها با هم را بررسی کنید و به صورت یک نمودار نمایش دهید. گروهی از داده های سالم که بیشترین همبستگی با نمونه های گروه بیمار دارند را از روی نمودار مشخص کنید تا در مراحل بعد از این گروه برای تحلیل های بعدی استفاده شود. به نظر شما لزوم انجام این مرحله چیست؟ (۳۰)