

پروژه نهایی مقدمه بیوانفورماتیک

هدف از این پروژه ، تجربه اولیه در تحلیل داده های بیوانفورماتیک می باشد. شما می بایست داده های Microarray مربوط به سرطان خون (لوکمیا) را تحلیل کنید. صفحه مربوط به دادگان و سایر اطلاعات مجموعه داده را می توانید در [لینک](#) مشاهده کنید.

لوکمیا انواع مختلفی دارد. یکی از انواع این سرطان ، لوکمی حاد مغز استخوان (AML) است. شما بایستی با تحلیل دادگانی که در اختیار دارید ، ژن هایی را که در این نوع سرطان نقش موثرتری دارند را به دست آورید. به این منظور در مجموعه دادگان ، داده هایی را که Phenotype آنها Normal است را گروه نرمال و نمونه هایی را که Source name آنها AML patient است را گروه تست در نظر بگیرید.

پایلین اجرایی:

● کنترل کیفیت داده (10 نمره):

ممکن است داده های اولیه آماده ی تحلیل نباشند و نیاز به تغییراتی داشته باشند. در صورت لزوم تغییراتی بر روی دادگان بدهید نمودار دادگان را پیش و پس از تغییراتی که داده اید رسم کنید تغییرات مذکور می توانند موارد مختلفی مثلاً نرمال سازی در صورت نیاز باشند

● کاهش ابعاد داده (10 نمره):

برخی ژن ها میزان بیان تقریباً برابری دارند. در این قسمت با استفاده از روش کاهش ابعاد ، تاثیر ژن هایی را که میزان بیان تقریباً برابر در نمونه ها دارند را کم کنید . داده هایی با بعد کمتر به دست آورید .

*امتیازی:

در این بخش در صورت استفاده از روش های کاهش ابعاد مختلف و مقایسه آن ها با یکدیگر و انتخاب خروجی بهترین روش برای مراحل بعد ، به شما نمره اضافه تعلق خواهد گرفت .

● بررسی همبستگی بین نمونه ها (10 نمره):

Source Name در نمونه های نرمال متفاوت است پس از کاهش ابعاد رابطه همبستگی بین نمونه ها را براساس Source Name بررسی کرده و خروجی را به شکل یک نمودار نمایش دهید

● بررسی تمایز در بیان ژن ها (20 نمره):

دسته نمونه های سالمی را که میزان همبستگی آن با نمونه ی تست بیشتر است را بیابید . سپس میزان بیان ژن ها را در دسته ی مذکور را با میزان بیان ژن ها در افراد بیمار مقایسه کنید . برخی از این ژن ها در افراد بیمار به صورت معنی داری بیان کمتری دارند و برخی دیگر به صورت معنی دار بیان بیشتری دارند . ژن های هر دسته را گزارش کنید (سطح معنی داری را 0.05 در نظر بگیرید).

● آنالیز gene anthology و pathway ها (20 نمره):

برای هر کدام از دسته ژن هایی که افزایش یا کاهش بیان در افراد بیمار دارند ، pathway ها و gene anthology مرتبط را گزارش کنید . (برای این کار می توانید از Enrichr کمک بگیرید.)

● موارد امتیازی دیگر :

- توضیح دهید موارد فوق چگونه در AML تاثیر گذارند.
- نتایج خود را در مقالات مرتبط در این زمینه مقایسه کنید .
- با جستجو در مقالات مختلف داروهایی که میتوانند در درمان این بیماری مفید باشند را جستجو کنید و مکانیزم آن ها را شرح دهید آیا مکانیزم آن ها با ژن هایی که شما پیدا کرده اید مرتبط می باشد؟
- هر مورد جالبی که به ذهن خودتان میرسد.

● گزارش پایانی (30 نمره) :

- گزارش پایانی نمره دارد. نتایج و تحلیل هایتان را به طور کامل توضیح دهید.
- موارد ارسالی باید شامل کدهای زده شده ،گزارش ،نمودار ها ، نتایج نهایی و هر تحلیل اضافه ای که خودتان انجام داده اید باشد.

● نکات پایانی :

- در تمام بخش ها برای تحلیل دادگان باید از زبان R استفاده کنید .(اگر از ایده های شبکه عصبی استفاده کردید برای بخش های نمره اضافه میتوانید از pytorch استفاده کنید ولی برای ایده های ماشین لرنینگ در صورت استفاده در بخش نمره اضافه حتما باید از R استفاده کنید.)
- برای آشنایی بیشتر با پایپ لاین اجرایی می توانید ویدیوهای مربوطه در درس مقدمه بیوانفورماتیک در مکتب خانه را مشاهده کنید.
- بخش های امتیازی متناسب با کیفیت آنها تا سقف 35 نمره خواهند داشت.
- سوالات خود را در کوئرا بپرسید.
- مهلت تحویل پروژه ساعت 23:59 روز 15 بهمن ماه است.