

به نام خدا

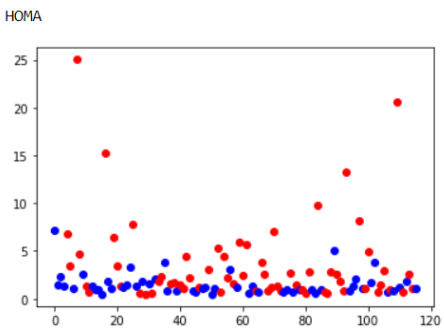
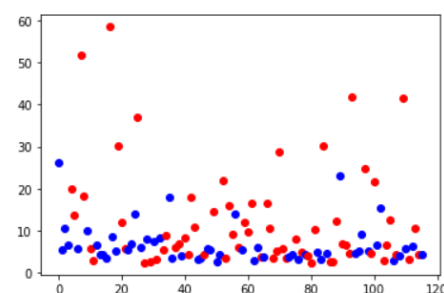
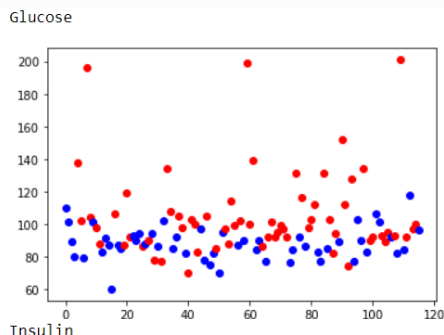
پروژه اول

Sarmeili.f@gmail.com

فراز سرمیلی

تحلیل نمودارها:

نمودارهای تک فیچر:



با توجه به اینکه برای هر فیچر یک نمودار رسم شده و دیتاپوینت‌ها بر اساس کلاس، رنگ خاص خود را دارند، میتوان گفت یکی از نمودارهایی که به شکل حدودی جداپذیر تر می‌باشد نمودار گلوکز است که در پزشکی یکی از اندیکاسیون‌های وجود تومور در بدن می‌باشد. به غیر از گلوکز به انسولین هم میتوانیم اشاره کنیم ولی باید دقت کنیم که انسولین هورمون طبیعی کاهش قند خون در بدن است و بدیهی است که با زیاد شدن قند انسولین نیاز زیاد شود. در نمودار دو فیچر گلوکز و انسولین رابطه مستقیم بین این دو پارامتر را به شکل واضح مشاهده میکنیم. فیچر HOMA یا Homeostatic model assessment فیچری است که به کمک آن عملکرد سلول‌های بتای پانکراس و به طور کلی معیاری است برای میزان مقاومت به انسولین و همانطور که قابل ملاحظه است به نظر میرسد برخی از افراد سرطانی در این فیچر نمره بیشتری را به نسبت دیگران کسب کرده اند و یعنی مقاومت به انسولین بیشتری هم نسبت به افراد غیر بیمار دارند که باز این فیچرها را به هم مربوط میکند. به طور کلی در پزشکی بالا رفتن میزان قند خون و مقاومت به انسولین از نشانه‌های پیشروی سرطان می‌باشد.

نمودارهای با دو فیچر:

در این دست از نمودارها، نمودارهایی که به طور کلی یک محور آن‌ها گلوکز یا پارامتر مربوط به آن است جداپذیرتر می‌باشند ولی به طور کلی با استاندارد اسکیل به علت وجود outlayer ها خیلی این جدا پذیری قابل تشخیص نبود به همین دلیل من سعی کردم از quantile transformation جهت پریپراسسینگ دیتاها استفاده کنم که همانطور که ملاحظه میفرمایید در نوتبوک تشخیص را مقداری ساده تر کرده است. یکی از فیچرهایی که شاید در سرطان حائز اهمیت باشد میزان پروتئینی است به نام رزیستین که از سلول‌های چربی به شکل عادی ترشح می‌شود و در شرایط التهابی میزان آن بیشتر میشود. پلات این فیچر به همراه گلوکز را در رو به رو میبینیم. همچنین رابطه‌ی بین ژن MCP1 و گلوکز نیز قابل ملاحظه می‌باشد که در مرز باریکی شاید بتوان با خطی آن‌ها را جدا کرد.

نتایج:

	Train Score	Test Score
Logistic Regression	0.760776	0.732609
KNN(K=1)	1.000000	0.672464
KNN(K=2)	0.829757	0.594565
KNN(K=3)	0.836255	0.664493
KNN(K=4)	0.749953	0.663043
KNN(K=5)	0.775923	0.706522
KNN(K=6)	0.749977	0.715942
KNN(K=7)	0.782305	0.672464
KNN(K=8)	0.769331	0.663043
KNN(K=9)	0.752057	0.672101
KNN(K=10)	0.771529	0.680797
SVM(kernel=linear)	0.754324	0.681159
SVM(kernel=rbf)	0.870734	0.785145
Decision Tree	1.000000	0.682246

با توجه به اکیورسی مدل‌ها به نظر می‌آید مدل SVM به کمک کرنل فانکشن RBF بهترین عملکرد را برای کلاسیفیکیشن دارد. عملکرد بقیه مدل‌ها نیز در جدول مقابل قابل ملاحظه می‌باشد:

