**چکیده :**

صورت مسیله پیاده سازی الگوریتم SVM یا Support Vector Machine روی تعدادی (241 نفر با 47 ویژگی ) است که برای آن از linear svm ,kernel svm استفاده کردیم ودر کرنل نیز از کرنل های مختلف استفاده کردیم و در هر حالت دقت الگوریتم را اندازه گیری کردیم

دیتاست از : <https://archive.ics.uci.edu/ml/machine-learning-databases/00489/>

تهیه شده است.

**مروری بر SVM :**

یکی های از روش های نظارتی برای پیش بینی مقادیر است یعنی داده ها دارای لیبل هستند و هدف آموزش الگوریتم است که بتواند روش کلاس بندی شدن n داده را یاد بگیرد تا بتواند کلاس داده ی n+1 را پیش بینی کند

در حل این مسیله 2 کلاس داریم کلاس 0 و کلاس 1 که کلاس صفر افراد دارای بیماری پارکینسون هستند و 1 افرادی که بیماری پارکینسون ندارند

در ساده ترین حالت svm داده ها خطی است و می توان خط مستقیم رسم کرد که داده هارا به 2 دسته ی مجزا تقسیم کند در این حالت از linear svm استفاده می کنیم

اما زمانی که داده هابه طور خطی قابل جداسازی نیستند یعنی نمی توان هیچ خط مستقیمی رسم کرد که با دقت بالایی داده ها را به 2 دسته تقسیم کرد در این حالت از kernel svm استفاده میکنیم که هدف آن ایجاد تابع تشابه است

مثلا داده هایی داریم که در 2 بعدی که می بینیم قابل جدا سازی نیستند اما اگر به آن ها یه بعد دیگر اضافه کنیم (که این کار با ضرب داخلی یا دات کردن انها قابل اعمال است )می توان آن هارا به طور خطی جدا کرد

در اینجا 4ازکرنل استفاده کردیم که دقت هرکدام از روش ها با توجه به مقادیر متفاوت داده ی تست و داده ی ترین در فایل ضمیمه وجود دارد

هدف svm فیت کردن بهینه ترین خطی است که می توان برای جدا سازی کلاس ها رسم کرد

اصطلاحات :

Support Vectors :

نزدیک ترین نقطه داده هایی که است که در کنار هایپرپلین یا سطح جداسازی وجود دارد و به ساختار طبقه بندی نزدیک تر است

Hyperplane :

سطح جداسازی که داده هارا جدا میکند

Margin :

گپ یا فضای خالی یا فاصله ی موجود بین نزدیک ترین نقطه از هر 2 کلاس

اطلاعات بیشتر درباره ی هر کرنل و عملکرد آن :

<https://www.datacamp.com/community/tutorials/svm-classification-scikit-learn-python>

<https://www.youtube.com/watch?v=mA5nwGoRAOo&feature=youtu.be>

ویدیوهای سری 20 تا 33

<https://jakevdp.github.io/PythonDataScienceHandbook/05.07-support-vector-machines.html>

**PCA**

Principle component anlysis

روشی برای تغییر داده های با ابعاد بالا به داد های با ابعاد پایین تر با استفاده از انتخاب ویژگی هایی که بیشتر توصیف از کلاس بندی داده هارا دارند که این انتخاب با استفاده از انتخاب داده هایی است که بیشترین واریانس را در خروجی ایجاد کرده اند

داده هارا با pca و بدون pca اعمال کردیم و دقت بااعمال pca 4 درصد بالا تر است.

**نتیجه گیری :**

پس از اعمال هر 4 کرنل و مقایسه دقت ها با یکدیگر نهایتا بیشتر دقت برای کرنل سیگمویید می باشد به این علت که تابع سیگمویید 2 عدد 0 و 1 را برمیگرداند پس برای کلاس بندی ها مناسب تراست

باتشکر

اعضا گروه

طهورا مروتی

مریم حجازی

زهرا بخشی