

تمرین سری ۳ واحد درسی داده کاوی مجازی

جناب آقای دکتر فراهانی و جناب آقای دکتر خردپیشه
دستیار آموزشی : علی شریفی

۲۶ اردیبهشت ۱۴۰۰

توجه کنید شما میتوانید بر روی کگل یا کولب و یا کامپیوتر های شخصی خود کار کنید .
به جای دانلود و آپلود دیتاست در گوگل درایو برای استفاده در کولب میتوانید به شیوه زیر عمل کنید .
چگونه از دیتاست های کگل در کولب استفاده کنیم ؟
ددلاین تمرین تا تاریخ ۷ خرداد می باشد .

۱ تمرین ها

۱. در خصوص کرنل های پرکاربرد روش SVM تحقیق کنید . به صورت کلی چرا ما از ایده کرنل در بحث SVM بهره میبریم . آیا میتوان در خصوص کرنل ها و استفاده ی آنها حکم کلی داد . به طور مثال بگوییم از کرنل RBF در این مواقع خاص استفاده میکنیم .
۲. قبلا با دیتاست کلاس بندی قیمت موبایل در **کگل** کار کرده ایم . بر روی دیتاست ، روش SVM را اجرا کنید . (استفاده از پکیج ها همانند sklearn مجاز است .)
۳. برای سوال ۲ حداقل ۵ حالت مختلف از قبیل کرنل ها و پارامترها را بررسی کنید و نتایج آن را گزارش دهید .
۴. برای سوال ۲ سعی کنید مبحث soft margin و hard margin را بررسی کنید و نتایج آن را گزارش دهید .
۵. مهندسی ویژگی یکی از بخش های مهم در فرایندهای علم داده میباشد . بر روی دیتاست موارد زیر را اجرا کنید .

- (آ) بر روی فیچر battery power از روش binning استفاده کنید . (حداقل سه اندازه مختلف برای بین ها در نظر بگیرید و حتی سائز بین ها را نامساوی در نظر بگیرید .)
- (ب) بر فیچرهای کتگوریکال در دیتاست one hot encoding را اعمال کنید . چرا ما باید به صورت کلی از این کدگذاری بهره ببریم .
- (ج) بررسی کنید آیا استفاده از تبدیل هایی از قبیل log transform و یا تبدیل نمایی در اینجا کاربرد دارد . به صورت کلی چرا از این دست تبدیلات بهره میبریم . (در این بخش شما مجاز هستید اگر تبدیل دیگری را مناسب میدانید اعمال کنید این بخش نمره امتیازی برای شما خواهد داشت . حتما دلیل استفاده از تبدیل استفاده شده را بیان کنید .)
- (د) یک فیچر جدید به نام مساحت یا حجم گوشی بسازید .
۶. برای هریک از حالت های سوال ۵ یک مدل SVM بسازید و بررسی کنید یکبار هم هر ۵ حالت را باهم اعمال کنید و مدل SVM روی آنها اجرا کنید . حاصل این مدل ها را گزارش کنید .