

تمرین سری ۳ واحد درسی یادگیری ماشین

جناب آقای دکتر فراهانی
دستیاران آموزشی : نوید کاشی ، علی شریفی

۳۱ اردیبهشت ۱۴۰۰

توجه کنید شما میتوانید بر روی کگل یا کولب و یا کامپیوترهای 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 روز بعد از امتحان درس یادگیری ماشین مقطع کارشناسی میباشد .

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 .
۷. در خصوص الگوریتم های مختلف ساخت درخت تصمیم (همانند ID3 ، CART و ...) تحقیق کنید . به صورت کلی تفاوت الگوریتم های مختلف ساخت درخت تصمیم در چیست ؟
۸. به دلخواه با استفاده از پکیج ها بر روی دیتاست مطرح شده یک درخت تصمیم بسازید .
۹. برای درخت تصمیم پارامتر های مختلف مورد ارزیابی قرار دهید . آیا عمق درخت و تعداد نمونه های موجود در هر هر گره تاثیری در عملکرد درخت تصمیم دارد ؟
۱۰. در خصوص هرس کردن Pruning درخت تصمیم تحقیق کنید . چرا ما به بحث هرس کردن درخت تصمیم نیاز دارد و چه کمکی به ما میکند .
۱۱. (بخش امتیازی) سعی کنید این هرس کردن درخت در مدل خود اجرا کنید و بررسی کنید آیا این هرس کردن در نتایج شما تاثیر داشته است .
۱۲. بر روی دیتاست یک random forest اجرا کنید و نتایج با یک درخت تصمیم مقایسه کنید . بررسی کنید آیا نتایج درخت تصمیم بهتر است و یا random forest و چرا ؟
۱۳. تحقیق کنید چرا با وجود روش های جدید از قبیل یادگیری عمیق و شبکه عصبی ، هم چنان روشی مانند درخت تصمیم محبوب است ؟ (راهنمایی : میتوانید در خصوص بحث تفسیری پذیری این مدل ها صحبت کنید .)

۱۴. (بخش امتیازی) در درخت های تصمیم ما قوانینی استخراج میکنیم و از این قوانین استفاده میکنیم. در خصوص روش های دیگری که به استخراج قوانین از روی دیتاست میپردازند (rule induction) همانند روش Ripper ، IREP ، ... تحقیق کنید و آن ها را توضیح دهید . (حداقل دو روش)