



دانشگاه صنعتی شریف- دانشکده برق

درس یادگیری ماشین

تمرین کامپیوتری سری اول

مهلت تحویل: 17 فروردین 1403

نیمسال دوم 1402-1403

آداب انجام تمرینات کامپیوتری

- **همفکری** دانشجویان در انجام تمرینات مانعی ندارد ولی تمرینات باید به صورت مشخص توسط خود تحویل دهنده انجام گرفته باشد. در زمان تصحیح تمرینات شباهت تمرینات بررسی خواهد شد و تمرین‌های با شباهت غیرعادی مورد قبول نیست.
- استفاده از منابع اینترنتی برای گرفتن ایده‌ی حل سوالات مانعی ندارد ولی کپی کردن پاسخ تمرینات از هر منبعی مورد قبول نیست.
- استفاده از کتابخانه‌های آماده در صورتی که در صورت تمرین قید نشده باشد، مانعی ندارد.
- کدهای تحویل شده باید قابلیت اجرای دوباره را داشته باشند. در صورت وجود مشکل در زمان تصحیح (حتی به دلیل خطای تایپی)، قسمت‌های غیرقابل اجرا مورد قبول نیست. لطفاً قبل از تحویل تمرین حتماً کد خود را بررسی فرمایید.
- فایل ZIP شامل کدهای تمرینات همراه گزارش (اگر در صورت تمرین نیاز به گزارش اعلام شده باشد) در CW آپلود گردد.

مسئله اول

- 1- در این سوال یک نمونه ۱۰۰ تایی داده در اختیار شما قرار گرفته است و از شما انتظار می رود که با استفاده از الگوریتم parzen-window با کرنل گوسی، توزیع آماری این نمونه ها را تخمین بزنید.
- 2- سپس تحلیل کنید که این نمونه داده ها از چه نوع توزیعی تولید شده اند.

مسئله دوم

- 1- در این سوال به شما یک دیتاست داده شده است که شما باید ابتدا آن را لود کرده و سپس داده های غیر عددی آن را به عدد تبدیل کرده و همین طور دیتاست را نرمالیزه کنید.
- 2- سپس با استفاده از الگوریتم Naïve_bayes، داده ها را طبقه بندی کنید.
- 3- سپس با استفاده از الگوریتم KNN، عملیات cross-validation را انجام داده و به ازای مقادیر مختلف K ، معیار های ذکر شده ارزیابی طبقه بند را بدست آورید و به کمک آن ها، منحنی ROC را برای این طبقه بند رسم کنید. سپس با استفاده از منحنی ای که بدست آوردید، مقدار بهینه K را پیدا کرده و به ازای آن ۵ مرتبه داده ها را به صورت رندوم به دو دسته آموزش و تست تقسیم کنید و میانگین معیار های ذکر شده ارزیابی طبقه بند را بدست آورید.
- 4- سپس همین کار را با الگوریتم KNN وزن دار تکرار کنید.
- 5- حال مینیمم و ماکزیمم معیار $F1_score$ برا این دو حالت از طبقه بند KNN را با یکدیگر مقایسه کرده و دلیل تفاوت آن ها را ذکر کنید.

مسئله سوم

- 1- در این سوال قرار است شما با استفاده از یک الگوریتم unsupervised و به کمک تخمین چگالی، یک گراند یک ویدیو را حذف کنید. ابتدا قسمت اول کد را ران کرده و سپس برای راحتی کار، فریم های ویدیو را با ضریب ۲ down_sample کنید.
 - 2- سپس میانگین و واریانس ها را برای هر پیکسل بدست آورید.
 - 3- در انتها با الگوریتم ذکر شده عملیات حذف یک گراند را انجام دهید و تاثیر پارامتر β را در دقت حذف یک گراند مشخص کنید.
- همراه با فایل ZIP کدها و خروجی، یک فایل گزارش شامل توضیحات در مورد آن چه انجام دادید و سوالات پرسش شده ارسال فرمایید. در صورت ارسال فایل نوتبوک می توانید توضیحات را در همان فایل بنویسید و نیازی به گزارش نیست.