



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)

به نام خدا

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

دانشکده برق

یادگیری ماشین – نیم سال اول 1402-1403

تمرین عملی سری چهارم درس مقدمه‌ای بر یادگیری ماشین: ماشین بردار پشتیبان

۱- در این سوال، هدف پیاده‌سازی الگوریتم SVM است که قادر به پیش‌بینی تصمیمات دانش‌آموزان در پایان دوره تحصیلی عادی باشد، از جمله اینکه آیا آن‌ها تحصیلات خود را ادامه می‌دهند، ترک تحصیل می‌کنند یا فارغ التحصیل می‌شوند. این پیش‌بینی بر اساس یک مجموعه داده‌ها (Prediction of students performance) ایجاد می‌شود که شامل اطلاعات مسیر تحصیلی، وضعیت جمعیت‌شناختی و عوامل اجتماعی-اقتصادی دانش‌آموزان در زمان ثبت‌نام می‌باشد.

الف) ابتدا مجموعه داده‌ها را به سه قسمت داده‌ها آموزش، اعتبارسنجی و تست با نسبت مناسب تقسیم کنید. سپس با استفاده از الگوریتم SVM خطی و مجموعه داده‌ها را آموزش دهید. صحت طبقه‌بندی را در مجموعه‌های آموزش و تست محاسبه کنید. سپس تعداد بردارهای پشتیبان را مشخص کنید.

ب) با استفاده از الگوریتم SVM غیرخطی در حالت soft SVM یک یادگیر SVM بسازید. پارامترهای مناسب یادگیر SVM را با استفاده از داده‌ها اعتبارسنجی بدست آورید تا بیشترین صحت را بروی مجموعه داده‌ها آن داشته باشید. میزان صحت طبقه‌بندی را روی مجموعه‌های آموزش و تست در این بهترین مقدار انتخابی محاسبه کنید. سپس تعداد بردارهای پشتیبان را مشخص نمایید و با قسمت قبل مقایسه کنید.

پ) این بار کرنل‌های چند جمله‌ای و RBF انتخاب و پارامترهای مناسب یادگیر SVM را با استفاده از داده‌ها اعتبارسنجی بدست آورید تا بیشترین صحت را بروی مجموعه داده‌ها آن داشته باشید. میزان صحت طبقه‌بندی

را روی مجموعه های آموزش و تست در این بهترین مقدار انتخابی محاسبه کنید. سپس تعداد بردارهای پشتیبان را مشخص نمایید و با قسمت‌های قبل مقایسه کنید.

ت) با الگوریتم SVM غیرخطی در حالت ترکیبی کرنل و soft SVM یک یادگیر SVM بسازید. کرنل را بهترین کرنل مرحله انتخاب کرده و میزان صحت طبقه‌بندی را روی مجموعه های آموزش و تست در این بهترین مقدار انتخابی محاسبه کنید. سپس تعداد بردارهای پشتیبان را مشخص نمایید و با قسمت‌های قبل مقایسه کنید.

ث) بخش قبل را با 3-fold cross validation تکرار و با قسمت‌های قبل مقایسه کنید.

۲- در بخش دوم تمرین، هدف پیاده‌سازی الگوریتم SVR خطی برای پیش‌بینی میزان تولید CO_2 تولید شده توسط یک خودرو است. بدین منظور با استفاده از مجموعه داده‌گان Fuel Consumption که در فایل پیوست موجود است، که شامل نوع خودرو، اطلاعات موتور و میزان مصرف سوخت خودرو در مسیرهای مختلف می‌باشد، این الگوریتم را آموزش دهید. برای پیاده‌سازی، مجموعه داده‌گان را به 3 قسمت Train، Test و Validation تقسیم نموده (با نسبت 60٪، 20٪ و 20٪) و سپس با استفاده از ویژگی‌های عددی این مجموعه داده، الگوریتم را آموزش دهید.

الف) برای SVR خطی، دقت مدل را برای داده‌های Train، Test و Validation به طور کامل گزارش کنید و بر روی نتایج بحث نمایید.

ب) (امتیازی) برای SVR غیرخطی، از بین کرنل‌های متداول، بهترین کرنل را معرفی نموده و دقت مدل را برای داده‌های Train، Test و Validation به طور کامل گزارش کنید و بر روی نتایج بحث نمایید.

تمرین عملی سری چهارم درس مقدمه‌ای بر یادگیری ماشین: ماشین بردار پشتیبان

گزارش: برنامه‌های نوشته شده در MATLAB یا پایتون به همراه نتایج و شکل‌های خواسته شده، تحلیل نتایج و نمودارها را ضمیمه کنید.

توجه: استفاده از کتابخانه‌ها در حل این سوال مجاز است.

توجه: کل تمرین و گزارش باید بصورت انفرادی نوشته شود و به برنامه‌ها و گزارش‌های کپی شده نمره‌ای تعلق نمی‌گیرد.

موعد تحویل: جمعه ۶ بهمن ۱۴۰۲

موعد نهایی تحویل با تاخیر: سه‌شنبه ۱۰ بهمن ۱۴۰۲ ساعت ۱۱ صبح

روش تحویل: آپلود در Courses

موفق و پیروز باشید