

پاییز ۱۳۹۸

یادگیری ماشین

تمرین سوم

پرهام الوانی

۹۸۱۳۱۹۱۰

۱. دسته‌بند بیز و رگرسیون لاجستیک هر دو برای دسته‌بندی داده‌هایی هستند که از نظر خطی جداپذیر می‌باشند. دسته‌بند رگرسیون لاجستیک دارای پارامتر وزن می‌باشد. این دسته‌بند در صورتی که وزن‌های آن به صورت مشخصی انتخاب شوند به دسته‌بند ساده گاوسی بیز تبدیل می‌شود.

۲. در دسته‌بند بیز ساده ممکن است برخی از حالت‌ها در داده‌های آموزش وجود نداشته باشند. عملیات هموارسازی به همه‌ی دسته‌ها یک مقدار ثابت اضافه می‌کند به این ترتیب این حالت‌ها از بین می‌روند. این روش به Laplace Smoothing معروف است.

۳.

۴. داده‌های شکل دو به صورت خطی جداپذیر هستند و داده‌ی مشخص شده از کلاس A است. داده‌های شکل یک از به صورت خطی جداپذیر نمی‌باشند بنابراین ابتدا با تبدیل می‌توانیم آن را به یک فضای جداپذیر خطی ببریم و با این ترتیب داده‌ی مشخص شده از کلاس B خواهد بود.

۵. در رگرسیون لاجستیک در صورتی که ابعاد داده‌ها زیاد باشد می‌توان با منظم‌سازی تابع هدف را برای وزن‌های بزرگ جریمه کرد. این منظم‌سازی از بیش‌برازش داده‌ها جلوگیری می‌کند و بنابراین احتمال تشخیص صحیح برچسب در داده‌های آموزش کاهش یافته است و این در حالی است که این کاهش در داده‌های آموزش بیشتر است.

۶. اضافه کردن ضریب منظم‌سازی در یک دسته‌بند از بیش‌برازش آن جلوگیری کرده و تاثیر نویز را روی آن کمتر می‌کند. هرچه ضریب منظم‌سازی در توابع هزینه بیشتر شود، احتمال بیش‌برازش کمتر می‌شود و ضرایب خروجی مدل کوچکتر می‌شوند. در صورتی که این وزن برای هر یک از ویژگی‌ها مقدار بزرگی باشد آن ویژگی در نظر گرفته نمی‌شود.