

تمرین دوم مبانی یادگیری ماشین (بخش 4)

محمد رضا ضیالاری (97222057)

- 1- یکی از راه ها این است که کلاس بندی مجدد انجام دهیم و آن ها را به دو کلاس (باینری) ببریم و رگرسیون را بر روی آن اجرا کنیم .
- 2- در حالتی که داده ها متوازن نبود نتایج بهتری به دست آمد که با توجه به حجم کم داده ها و تراکم شدید داده ها در یک کلاس این اتفاق مورد انتظار بود .
- 3- بله دقت در حدود 20 درصد بعد از feature selection رشد پیدا کرد . زیرا از ویژگی های مهم و تاثیر گذار تر استفاده شد .
- 4- الف) روش توکار: انتخاب ویژگی را در فرآیند آموزش انجام می دهند و معمولا برای ماشین های یادگیری خاصی مورد استفاده قرار می گیرند. در این روش ها، جست و جو برای یک زیرمجموعه بهینه از ویژگی ها در مرحله ساخت دسته بندی انجام می شود و می توان آن را به عنوان جست و جویی در فضای ترکیبی از زیر مجموعه ها و فرضیه ها دید. این روش ها قادر به ثبت وابستگی ها با هزینه های محاسباتی پایین تر نسبت به بسته بندی ها هستند.
- ب) روش حذف بازگشتی ویژگی : در این روش ویژگی ها به طور بازگشتی و با در نظر گرفتن مجموعه های کوچک و کوچک تر از ویژگی ها انتخاب میشوند. در این روش ویژگی ها بر اساس مرتبه حذف شدن آن ها از فضای ویژگی رتبه بندی می شوند. ابتدا بر روی یک مجموعه اولیه از ویژگی ها آموزش داده میشود و سپس با استفاده از معیارهای مختلف مانند ضریب و یا مهم بودن فیچر ها، فیچرهای خاصی انتخاب میشوند و تعداد خاصی از کل را انتخاب میکنیم. (منبع فرادرس)

5- در هر دو نمی توان به طور قطع اعلام کرد که ویژگی های انتخاب شده ، تاثیرگذارترین ویژگی ها هستند . یکی از راهکار ها این است که بطور ترکیبی از هر دو برای استخراج و انتخاب تاثیرگذارترین ویژگی ها استفاده کنیم .

6 - این روش نیز مانند روش PCA ابعاد را برای ما کاهش می دهد . که از طریق projection این کار را انجام می دهد . در مسائلی که داده ها پراکندگی بسیاری دارند ، استفاده از LDA نسبت به PCA برتری دارد .

8- یک روش برای کلاس بندی باینری است که خروجی آن یک عدد (1یا1یا0) می باشد .
1 برای زمانی است که پیش بینی ما به واقعیت نزدیک است . -1 برای زمانی است که پیش بینی ما مخالف واقعیت است و 0 برای زمانی است که پیش بینی دقیقی نداشتیم .