بررسی تاثیر تعداد فیچر‌ها بر دقت مدل

# مدل adaboost

## انتخاب با معیار mutual information

در این جا اگر نیازمند l فیچر باشیم آن l فیچری را برمی‌داریم که بیشترین میزان mutal information با فیچر لیبل داده‌ها را دارند. میزان دقت مدل آدابوست با این روش انتخاب ویژگی در جدول زیر آمده است:

|  |  |
| --- | --- |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.48 |
| 10 | 0.50 |
| 50 | 0.82 |
| 100 | 0.84 |
| 561 | 0.86 |

## انتخاب با استفاده از SVM پنالتی‌دار

در این روش بر این نحو عمل میکنیم که مانند بخش دوم، مدل svm پنالتی‌داری را روی داده‌ها آموزش می‌دهیم، سپس آن فیچرهایی که در آن svm مقدارشان کمتر از سایر فیچر‌ها شده و یا نزدیک به صفر هستند را مورد صرف‌نظر قرار می‌دهیم. نتایج دقت مدل حاصل از این روش در جدول زیر آمده است:

|  |  |
| --- | --- |
| Svm با پنالتی l1 و ضریب منظم سازی 0.01 | |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.65 |
| 10 | 0.82 |
| 50 | 0.88 |
| 100 | 0.87 |
| 561 | 0.83 |

لازم به ذکر است که svm استفاده شده توسط ما با پارامتر‌های پنالتی از نوع l1 و ضریب منظم‌سازی 0.01 بوده است.

# مدل logistic regression

## انتخاب با معیار mutual information

|  |  |
| --- | --- |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.59 |
| 10 | 0.63 |
| 50 | 0.74 |
| 100 | 0.93 |
| 561 | 0.96 |

## انتخاب با استفاده از SVM پنالتی‌دار

در ابتدا با استفاده از svm با پنالتی l1 و ضریب منظم‌سازی 0.01 عمل انتخاب ویژگی را انجام می‌دهیم، نتایج این امر در جدول زیر آمده است:

|  |  |
| --- | --- |
| Svm با پنالتی l1 و ضریب منظم سازی 0.01 | |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.76 |
| 10 | 0.85 |
| 50 | 0.95 |
| 100 | 0.96 |
| 561 | 0.96 |

سپس برای بررسی بیشتر این امر را با یک svm با پنالتی l2 و ضریب منظم‌سازی 0.01 نیز تکرار می‌کنیم، برای نتایج داریم:

|  |  |
| --- | --- |
| Svm با پنالتی l2 و ضریب منظم سازی 0.01 | |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.72 |
| 10 | 0.83 |
| 50 | 0.95 |
| 100 | 0.96 |
| 561 | 0.96 |

# مدل neural network

## انتخاب با معیار mutual information

|  |  |
| --- | --- |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.56 |
| 10 | 0.56 |
| 50 | 0.86 |
| 100 | 0.88 |
| 561 | 0.96 |

## انتخاب با استفاده از SVM پنالتی‌دار

|  |  |
| --- | --- |
| Svm با پنالتی l1 و ضریب منظم سازی 0.01 | |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.75 |
| 10 | 0.90 |
| 50 | 0.95 |
| 100 | 0.96 |
| 561 | 0.96 |

لازم به ذکر است که svm استفاده شده توسط ما با پارامتر‌های پنالتی از نوع l1 و ضریب منظم‌سازی 0.01 بوده است.

# مدل random forest

## انتخاب با معیار mutual information

|  |  |
| --- | --- |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.64 |
| 10 | 0.67 |
| 50 | 0.78 |
| 100 | 0.94 |
| 561 | 0.96 |

## انتخاب با استفاده از random forest

در این جا ابتدا مدل random forest با ابرپارامتر‌های تعیین شده را روی داده‌ها آموزش می‌دهیم و سپس هنگام انتخاب ویژگی از فیچر‌هایی که اهمیت کمتری در این مدل‌ داشته‌اند صرف‌نظر می‌کنیم:

|  |  |
| --- | --- |
| Random forest | |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.82 |
| 10 | 0.83 |
| 50 | 0.94 |
| 100 | 0.95 |
| 561 | 0.95 |

# مدل SVM

## انتخاب با معیار mutual information

|  |  |
| --- | --- |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.59 |
| 10 | 0.56 |
| 50 | 0.72 |
| 100 | 0.92 |
| 561 | 0.97 |

## انتخاب با استفاده از SVM پنالتی‌دار

در ابتدا با استفاده از svm با پنالتی l1 و ضریب منظم‌سازی 0.01 عمل انتخاب ویژگی را انجام می‌دهیم، نتایج این امر در جدول زیر آمده است:

|  |  |
| --- | --- |
| Svm با پنالتی l1 و ضریب منظم سازی 0.01 | |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.77 |
| 10 | 0.89 |
| 50 | 0.95 |
| 100 | 0.95 |
| 561 | 0.96 |

سپس برای بررسی بیشتر این امر را با یک svm با پنالتی l2 و ضریب منظم‌سازی 0.01 نیز تکرار می‌کنیم، برای نتایج داریم:

|  |  |
| --- | --- |
| Svm با پنالتی l2 و ضریب منظم سازی 10 | |
| تعداد ویژگی انتخاب‌شده | دقت مدل |
| 5 | 0.64 |
| 10 | 0.80 |
| 50 | 0.95 |
| 100 | 0.96 |
| 561 | 0.97 |