



Sign up

 Watch

1

★ Star

0

 Fork

0

Issues 0

Pull requests 0

Projects 0

Insights

Dismiss

Sign up

Find file

Copy path



362098c 17 minutes ago

1 contributor

Raw

Blame

History



1. Number of times pregnant
2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3. Diastolic blood pressure (mm Hg)
4. Triceps skin fold thickness (mm)
5. 2-Hour serum insulin (μ U/ml)
6. Body mass index ($\text{weight in kg}/(\text{height in m})^2$)
7. Diabetes pedigree function
8. Age (years)
9. Class variable (0 or 1)

داده های این پروژه مربوط به تشخیص دیابت در حدود 700 نفر است و هدف نهایی کلاس بندی داده ها به دو دسته ی سالم و بیمار می و decision tree و logistic regression ، LDA ، QDA ، gnb ، SVM ، random forest ، bagging و بدون پیش PCA در 4 حالت پیش پردازش خودرمنگار با دولايه پنهان ، پیش پردازش خودرمنگار با سه لایه پنهان ، پیش پردازش پردازش برای هرکدام بررسی نموده ایم و نتایج را با صورت های متنوع مقایسه کرده ایم

برای مشاهده کامل نتایج بدست آمده می‌توانید به لینک مقابل مراجعه کنید **نتایج بدست آمده برای پروژه**

Logistic regression

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.6315

4 , 2	0.5319
4 , 3	0.5132
5 , 2	0.5916
5 , 3	0.5545
5 , 4	0.6315
6 , 3	0.6818
6 , 4	0.5192
6 , 5	6154
7 , 2	0.5319
7 , 3	0.7078
7 , 4	0.5397
7 , 5	0.6813
7 , 6	0.5949

neurons	AUC
4 , 3 , 2	0.5
5 , 3 , 2	0.5192
5 , 4 , 2	0.4953
5 , 4 , 3	0.5975
6 , 3 , 2	0.5086
6 , 4 , 2	0.5272
6 , 5 , 2	0.5
6 , 5 , 3	0.4953
6 , 5 , 4	0.6069
7 , 4 , 2	0.5272
7 , 4 , 3	0.5358
7 , 5 , 2	0.5119
7 , 5 , 3	0.6639
7 , 5 , 4	0.6175
7 , 6 , 2	0.5179
7 , 6 , 3	0.5949
7 , 6 , 4	0.6473
7 , 6 , 5	0.6873

PCA pre-processing results

Components	AUC
1	0.5272

2	0.7099
3	0.7158
4	0.7112
5	0.7052
6	0.7112
7	0.7065

Without pre-processing

AUC	0.7665
-----	--------

Conclusion

در پیش پردازش با خودرمزنگار با دولایه logistic regression با توجه به مقادیر بدست آمده ، میتوان مشاهده کرد که بهترین حالت الگوریتم برابر با 0.7078 است. در حالت پیش پردازش با خودرمزنگار با سه لایه پنهان بهترین حالت با تعداد AUC پنهان در حالت 7 و 3 نوروں با مقدار بهترین حالت با 3 مولفه اصلی بدست می آید که مقدار آن برابر PCA 7 و 6 و 5 نوروں با مقدار 0.6873 است. در حالت پیش پردازش با در حالت بدون پیش پردازش برابر با 0.7665 میشود. باتوجه به مقادیر و مقایسه آن ها میتوان نتیجه AUC 0.7158 است. همچنین مقدار و در حالت پیش پردازش با خودرمزنگار با سه لایه AUC در حالت بدون پیش پردازش دارای بهترین مقدار logistic regression گرفت که بهتر است بدون logistic regression است که برابر با 0.4953 می باشد. بنابراین الگوریتم AUC پنهان با 6 و 5 و 3 نوروں دارای کمترین مقدار پردازش اجرا شود.

LDA

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.6222
4 , 2	0.5319
4 , 3	0.5132
5 , 2	0.6009
5 , 3	0.5545
5 , 4	0.6362
6 , 3	0.6652
6 , 4	0.5192
6 , 5	0.6108
7 , 2	0.5272
7 , 3	0.7078
7 , 4	0.5443
7 , 5	0.6813
7 , 6	0.6056

neurons	AUC
4 , 3 , 2	0.5

5 , 3 , 2	0.5192
5 , 4 , 2	0.4953
5 , 4 , 3	0.5835
6 , 3 , 2	0.5086
6 , 4 , 2	0.5272
6 , 5 , 2	0.5106
6 , 5 , 3	0.4953
6 , 5 , 4	0.6209
7 , 4 , 2	0.5272
7 , 4 , 3	0.5252
7 , 5 , 2	0.5119
7 , 5 , 3	0.6639
7 , 5 , 4	0.6175
7 , 6 , 2	0.5179
7 , 6 , 3	0.5996
7 , 6 , 4	0.6473
7 , 6 , 5	0.6873

PCA pre-processing results

Components	AUC
1	0.522
2	0.6992
3	0.7052
4	0.7005
5	0.7052
6	0.7065
7	0.6959

Without pre-processing

AUC	0.7665
-----	--------

Conclusion

در پیش پردازش خودرمنگار با دولایه پنهان مربوط به حالت AUC میتوان مشاهده کرد که بهترین مقدار LDA در بررسی و 3 نوروں است که مقداری برابر با 0.7078 دارد. در پیش پردازش خودرمنگار با سه لایه پنهان بهترین مقدار برابر با 0.6873 است که 7 بهترین حالت با 6 مولفه اصلی بدست می آید که مقداری برابر با PCA مربوط به تعداد 7 و 6 و 5 نوروں می باشد. درحالت پیش پردازش با LDA در حالت بدون پیش پردازش برابر با 0.7665 است. به استناد اطلاعات بدست آمده میتوان گفت که AUC 0.7065 دارد. همچنین مقدار است بنابراین بهتر است این الگوریتم بدون پیش پردازش اجرا شود و همچنین AUC در حالت بدون پیش پردازش دارای مقدار مطلوب تر درحالت پیش پردازش خودرمنگار با سه لایه پنهان و تعداد 5 و 4 و 2 نوروں و همچنین 6 و 5 و 3 نوروں است که مقداری AUC کمترین مقدار برابر با 0.4953 دارد.

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.6328
4 , 2	0.5192
4 , 3	0.5537
5 , 2	0.569
5 , 3	0.5
5 , 4	0.5
6 , 3	0.6499
6 , 4	0.5205
6 , 5	0.5
7 , 2	0.5
7 , 3	0.6652
7 , 4	0.5171
7 , 5	0.666
7 , 6	0.5

neurons	AUC
4 , 3 , 2	0.5
5 , 3 , 2	0.5
5 , 4 , 2	0.5
5 , 4 , 3	0.5
6 , 3 , 2	0.5205
6 , 4 , 2	0.5
6 , 5 , 2	0.5145
6 , 5 , 3	0.5
6 , 5 , 4	0.5929
7 , 4 , 2	0.5
7 , 4 , 3	0.5
7 , 5 , 2	0.5418
7 , 5 , 3	0.5996
7 , 5 , 4	0.5
7 , 6 , 2	0.4927
7 , 6 , 3	0.5
7 , 6 , 4	0.5

7 , 6 , 5

0.7052

PCA pre-processing results

Components	AUC
1	0.5358
2	0.7086
3	0.7299
4	0.6852
5	0.6806
6	0.6912
7	0.7005

Without pre-processing

AUC

0.7431

Conclusion

است که برابر با 0.666 است. در AUC با پیش پردازش خودرمنگار با دولایه پنهان حالت 7 و 5 نوروں دارای بیشترین مقدار QDA در بررسی نیز PCA پیش پردازش خودرمنگار با سه لایه پنهان مقدار 0.7052 بیشترین است که مربوط 7 و 6 و 5 نوروں است. در حالت پیش پردازش برابر با AUC آن برابر با 0.7299 است. در حالت بدون پیش پردازش مقدار AUC بیشترین مقدار به حالت 3 مولفه تعلق دارد که مقدار 0.7431 شده است. با توجه به مقادیر بدست آمده به این نتیجه میرسیم که حالت بدون پیش پردازش دارای بهترین مقدار است بنابراین نیز مربوط به پیش پردازش خودرمنگار با سه لایه پنهان و تعداد نوروں AUC بدون پیش پردازش بهتر است و کمترین مقدار QDA اجرای 7 و 6 و 2 است که مقداری برابر با 0.4927 دارد

GNB

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.6128
4 , 2	0.5264
4 , 3	0.5252
5 , 2	0.5882
5 , 3	0.5451
5 , 4	0.6328
6 , 3	0.6559
6 , 4	0.5086
6 , 5	0.5511
7 , 2	0.5405
7 , 3	0.6175
7 , 4	0.509

7 , 5	0.5537
7 , 6	0.5

neurons	AUC
4 , 3 , 2	0.486
5 , 3 , 2	0.5252
5 , 4 , 2	0.4953
5 , 4 , 3	0.5975
6 , 3 , 2	0.5205
6 , 4 , 2	0.5039
6 , 5 , 2	0.5145
6 , 5 , 3	0.4919
6 , 5 , 4	0.6388
7 , 4 , 2	0.5252
7 , 4 , 3	0.5205
7 , 5 , 2	0.5418
7 , 5 , 3	0.6665
7 , 5 , 4	0.5319
7 , 6 , 2	0.4927
7 , 6 , 3	0.5443
7 , 6 , 4	0.5895
7 , 6 , 5	0.7099

PCA pre-processing results

Components	AUC
1	0.5358
2	0.7086
3	0.7252
4	0.6992
5	0.6946
6	0.7112
7	0.7265

Without pre-processing

AUC	0.7431
-----	--------

Conclusion

مربوط به حالت 6و3 نوروں است کہ مقدار آن برابر باAUCدر حالت پیش پردازش خودرمزنگار با دولایہ پنهان بہترین مقدار gnb در حالت7و6 بیشترین است کہ برابر با0.7099 می AUC0.6559 می باشد. در حالت پیش پردازش خودرمزنگار با سه لایہ پنهان مقدار دارای بیشترین مقدار برابر با 0.7265 می باشد کہ مربوط بہ 7مولفہ اصلی است. همچنین PCA درحالت پیش پردازش Gnb باشد در حالت بدون پیش پردازش دارای بیشترین gnb درحالت بدون پیش پردازش برابر با 0.7431 است. بنابراین میتوان گفت کہ AUCمقدار مربوط بہ حالت پیش پردازش خودرمزنگار با سه AUCاست و درنتیجہ دراین حالت مطلوب تر است و از طرفی کمترین مقدار AUCمقدار لایہ پنهان و تعداد 4و3و2 نوروں برابر با 0.486 است.

SVM

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.6103
4 , 2	0.5
4 , 3	0.5
5 , 2	0.5877
5 , 3	0.5272
5 , 4	0.6149
6 , 3	0.6806
6 , 4	0.5
6 , 5	0.6009
7 , 2	0.5
7 , 3	0.7065
7 , 4	0.4886
7 , 5	0.6967
7 , 6	0.5272

4 , 3 , 2	0.5
5 , 3 , 2	0.5319
5 , 4 , 2	0.4953
5 , 4 , 3	0.5392
6 , 3 , 2	0.5
6 , 4 , 2	0.5
6 , 5 , 2	0.5
6 , 5 , 3	0.5
6 , 5 , 4	0.5817
7 , 4 , 2	0.5
7 , 4 , 3	0.5
7 , 5 , 2	0.5119
7 , 5 , 3	0.6686

7 , 5 , 4	0.6136
7 , 6 , 2	0.5
7 , 6 , 3	0.5711
7 , 6 , 4	0.6461
7 , 6 , 5	0.692

PCA pre-processing results

Components	AUC
1	0.5
2	0.6886
3	0.7146
4	0.6946
5	0.6899
6	0.7158
7	0.6899

Without pre-processing

AUC	0.7618
-----	--------

Conclusion

آن می باشد. AUC با پیش پردازش خودرمنگار با دولایه پنهان حالت 7و3 نوروں درلایه ها است که مقدار SVM بهترین حالت در پیش SVM. برابر با 0.695 بهترین پاسخ است AUC درحالت پیش پردازش خودرمنگار با سه لایه پنهان حالت 7و6و5 نوروں با مقدار برابر با 0.7158 که با 6مولفه اصلی بدست می آید. در حالت بدون پیش پردازش AUC در بهترین حالت دارای مقدار PCA پردازش در حالت بدون پیش پردازش دارای بهترین مقدار SVM برابر با 0.7618 میشود. با توجه به اطلاعات بدست آمده میتوان گفت که AUC مقدار است که مطلوب تر است و همچنین در حالت پیش پردازش خودرمنگار با دو لایه پنهان و تعداد نوروں برابر با 7و4 دارای کمترین می باشد که برابر با 0.4886 است AUC مقدار

Random forest

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5432
4 , 2	0.5137
4 , 3	0.5215
5 , 2	0.6113
5 , 3	0.535
5 , 4	0.6211
6 , 3	0.5934
6 , 4	0.5487

6 , 5	0.6367
7 , 2	0.5339
7 , 3	0.638
7 , 4	0.5661
7 , 5	0.6772
7 , 6	0.5614

neurons	AUC
4 , 3 , 2	0.4844
5 , 3 , 2	0.5972
5 , 4 , 2	0.5
5 , 4 , 3	0.4841
6 , 3 , 2	0.6517
6 , 4 , 2	0.5679
6 , 5 , 2	0.5584
6 , 5 , 3	0.5013
6 , 5 , 4	0.681
7 , 4 , 2	0.5858
7 , 4 , 3	0.5249
7 , 5 , 2	0.5495
7 , 5 , 3	0.6639
7 , 5 , 4	0.5041
7 , 6 , 2	0.6079
7 , 6 , 3	0.5233
7 , 6 , 4	0.6652
7 , 6 , 5	0.6261

PCA pre-processing results

Components	AUC
1	0.4735
2	0.6614
3	0.7291
4	0.698
5	0.6886
6	0.7052
7	0.7418

Without pre-processing

AUC	0.7133
-----	--------

Conclusion

برابر 0.6772 AUC با پیش پردازش خودرمنگار با دولایه پنهان حالت 7 و 5 نورون در لایه ها با مقدار random forest بهترین حالت در روش در آن 0.681 است. در حالت AUC است. در پیش پردازش خودرمنگار با سه لایه پنهان حالت با 6 و 5 و 4 نورون بهترین حالت است که مقدار برابر با 0.7418 است. در حالت بدون پیش پردازش AUC بهترین حالت در زمان است که 7 مولفه اصلی داریم که مقدار PCA پیش پردازش برابر با 0.7418 random forest برای الگوریتم AUC مقدار برابر با 0.7133 است. با توجه به مقادیر میتوان نتیجه گرفت که بهترین مقدار می باشد در نتیجه میتوان گفت اجرای الگوریتم با این روش بهتر است و کمترین مقدار نیز مربوط به PCA است که مربوط به پیش پردازش آن 0.4841 است AUC حالت پیش پردازش خودرمنگار با سه لایه پنهان که 5 و 4 و 3 نورون دارد و مقدار

Bagging

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.536
4 , 2	0.4878
4 , 3	0.5108
5 , 2	0.6206
5 , 3	0.521
5 , 4	0.5679
6 , 3	0.6279
6 , 4	0.5687
6 , 5	0.658
7 , 2	0.5092
7 , 3	0.6466
7 , 4	0.6266
7 , 5	0.6818
7 , 6	0.5402

neurons	AUC
4 , 3 , 2	0.475
5 , 3 , 2	0.6585
5 , 4 , 2	0.5
5 , 4 , 3	0.4688
6 , 3 , 2	0.6198
6 , 4 , 2	0.5679
6 , 5 , 2	0.529
6 , 5 , 3	0.5013

6 , 5 , 4	0.6372
7 , 4 , 2	0.5692
7 , 4 , 3	0.5806
7 , 5 , 2	0.5542
7 , 5 , 3	0.638
7 , 5 , 4	0.5181
7 , 6 , 2	0.6125
7 , 6 , 3	0.5632
7 , 6 , 4	0.6959
7 , 6 , 5	0.6533

PCA pre-processing results

Components	AUC
1	0.4556
2	0.7179
3	0.7052
4	0.6712
5	0.7065
6	0.7031
7	0.6759

Without pre-processing

AUC	0.7524
-----	--------

Conclusion

مربوط به حالت 7و5 نوروں است که مقدار آنAUCدر حالت پیش پردازش خودرمزنگار با دولایه پنهان بهترین مقدار bagging در روش برابر با 0.6818 می باشد. در پیش پردازش خودرمزنگار با سه لایه پنهان مقدار 0.6959 بیشترین است که مربوط به تعداد 6و4 نوروں بهترین حالت با23 مولفه اصلی بدست می آید که مقدار آن برابر 0.7179 است. در حالت بدون پیش PCA است. درحالت پیش پردازش با در حالت bagging برابر با 0.7524 بدست آمده است. با توجه به نتایج میتوان گفت که بهترین حالت برای الگوریتم AUC پردازش نیز مقدار بدون پیش پردازش بدست می آید و کمترین مقدار نیز در حالت پیش پردازش خودرمزنگار با سه لایه پنهان و تعداد نوروں 5و4و3 است که برابر با 0.4688 شده استAUC مقدار

Decision Tree

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5466
4 , 2	0.515
4 , 3	0.5474

5 , 2	0.6299
5 , 3	0.5303
5 , 4	0.6151
6 , 3	0.6006
6 , 4	0.5215
6 , 5	0.6533
7 , 2	0.5113
7 , 3	0.6027
7 , 4	0.5934
7 , 5	0.6852
7 , 6	0.6061

neurons	AUC
4 , 3 , 2	0.4844
5 , 3 , 2	0.6572
5 , 4 , 2	0.5
5 , 4 , 3	0.4854
6 , 3 , 2	0.5845
6 , 4 , 2	0.542
6 , 5 , 2	0.5503
6 , 5 , 3	0.5013
6 , 5 , 4	0.6164
7 , 4 , 2	0.5985
7 , 4 , 3	0.5474
7 , 5 , 2	0.5589
7 , 5 , 3	0.6712
7 , 5 , 4	0.4981
7 , 6 , 2	0.6125
7 , 6 , 3	0.5985
7 , 6 , 4	0.6619
7 , 6 , 5	0.6346

PCA pre-processing results

*	AUC
1	0.4807
2	0.6754
3	0.7179

4	0.6972
5	0.701
6	0.7151
7	0.7252

Without pre-processing

AUC	0.7511
-----	--------

Conclusion

است که برابر با AUC 0.6852 با پیش پردازش خودرمنگار با دولایه پنهان حالت 7 و 5 نوروں دارای بیشترین مقدار decision tree در بررسی است. در حالت پیش پردازش با خودرمنگار با سه لایه پنهان بهترین حالت با تعداد 7 و 5 و 3 نوروں با مقدار 0.6712 است. در حالت پیش بهترین حالت با 7 مولفه اصلی بدست می آید که مقداری برابر با 0.7252 دارد. در حالت بدون پیش پردازش نیز مقدار PCA پردازش با حالت بدون پیش پردازش decision tree برابر با 0.7511 بدست آمده است. بنابراین میتوان گفت که بهترین حالت برای اجرا الگوریتم AUC برابر با AUC نیز مربوط به حالت پیش پردازش با سه لایه پنهان و 4 و 3 و 2 نوروں است که مقدار AUC است و از طرفی کمترین مقدار دارد 0.4844.

Compared Models based on pre-processing methods

Auto encoder (with two hidden layers)

در LDA و logistic regression با بررسی نتایج مختلف در حالت پیش پردازش خودرمنگار با دولایه پنهان متوجه میشویم که روش در حالت 4 و 2 نوروں دارای bagging است که برابر با 0.7078 می باشد و همچنین روش AUC حالت 7 و 3 نوروں دارای بیشترین مقدار است که برابر با 0.4878 می باشد AUC کمترین مقدار

Auto encoder (with three hidden layers)

برابر با 0.7099 است که بزرگترین مقدار موجود در این AUC در حالت 7 و 6 و 5 نوروں دارای gnb با مشاهده نتایج بدست آمده میتوان گفت که در حالت 5 و 3 و 4 کمترین مقدار را دارد که برابر با bagging مربوط به روش AUC حالت می باشد و مطلوب ترین می باشد، همچنین است 0.4688.

PCA

با تعداد مولفه اصلی 1 تا 7 مشاهده میکنیم که بیشترین PCA با بررسی نتایج بدست آمده در روش های مختلف در حالت پیش پردازش random forest با 7 مولفه اصلی می باشد یعنی در این حالت از پیش پردازش random forest است که مربوط به 0.7418 ، AUC مقدار را دارد که برابر با 0.4556 است و در حالت 1 مولفه اصلی AUC کمترین مقدار bagging مطلوب تر است. همچنین میتوان مشاهده کرد که بدست می آید.

Without pre-processing

LDA و logistic regression مربوط به روش های AUC با توجه به نتایج ، میتوان گفت که در بررسی روش ها در حالت بدون پردازش بهترین مقدار است که برابر با 0.7133 می باشد random forest است که برابر با 0.7665 می باشد. همچنین کمترین مقدار نیز مربوط به regression

Conclusion

در حالت بدون پیش پردازش و در روش AUC با بررسی تمامی اطلاعات بدست آمده از کلاسیندی داده ها مشاهده میکنیم که بهترین مقدار بدست آمده است ، بنابراین کلاسیندی داده های این پروژه با این دو روش مطلوب تر است LDA و logistic regression

References

- [autoencoder-keras-tutorial](#)
- <https://keras.io/getting-started/sequential-model-guide/#stacked-lstm-for-sequence-classification>
- [Applied Deep Learning - Part 3: Autoencoders](#)

