



Why GitHub?

Enterprise

Explore

Marketplace

Pricing

Search



Sign in

Sign up

mohammadranjbar / data-mining-homeworks

Watch

1

Star

2

Fork

0

Code

Issues 0

Pull requests 0

Projects 0

Insights

Branch: master ▾

data-mining-homeworks / final-project /

Create new file

Find file

History



mohammadranjbar Update readme with new result of auto encoders

Latest commit c3f355b Feb 13, 2019

..

results	Set epoches 2000 for auto encoder	Feb 7, 2019
Readme.md	Update readme with new result of auto encoders	Feb 13, 2019
main.py	Set epoches 2000 for auto encoder	Feb 7, 2019
pima-indians-diabetes.csv	Calculate classification for pca and normal data	Jan 20, 2019

Readme.md

Preparing data

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 (mu U/ml)
6. Body mass index (weight in kg/(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 bagging و random forest ، SVM ، gnb ، QDA ، LDA ، logistic regression باشد.در این پروژه ما مسئله را با 8 روش و بدون پیش پردازش PCA در 4 حالت پیش پردازش خودرمننگار با دولایه پنهان ، پیش پردازش خودرمننگار با سه لایه پنهان ، پیش 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](#)

Analysis

Logistic regression

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5516
4 , 2	0.7205
4 , 3	0.6992
5 , 2	0.5
5 , 3	0.7099
5 , 4	0.6315
6 , 3	0.5192
6 , 4	0.7358
6 , 5	0.7312
7 , 2	0.7052
7 , 3	0.6899
7 , 4	0.6946
7 , 5	0.6959
7 , 6	0.7325

neurons	AUC
5 , 3 , 2	0.7112
5 , 4 , 2	0.5464
5 , 4 , 3	0.7052
6 , 3 , 2	0.5272
6 , 4 , 2	0.7158
6 , 5 , 2	0.5975
6 , 5 , 3	0.5272
6 , 5 , 4	0.7052
7 , 4 , 2	0.5272
7 , 4 , 3	0.5358
7 , 5 , 2	0.5013
7 , 5 , 3	0.6946
7 , 5 , 4	0.7265
7 , 6 , 2	0.5226
7 , 6 , 3	0.7205
7 , 6 , 4	0.7099
7 , 6 , 5	0.583

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 خودرمنزنگار با دولایه logistic regression با توجه به مقادیر بدست آمده ، میتوان مشاهده کرد که بهترین حالت الگوریتم برابر با 0.7358 است. در حالت پیش پردازش با خودرمنزنگار با سه لایه پنهان بهترین حالت با تعداد AUC پنهان در حالت 6 و 4 نوروں با مقدار بهترین حالت با 3 مولفه اصلی بدست می آید که مقدار آن برابر PCA 0.7158 و 7 و 4 نوروں با مقدار 0.7265 است. در حالت پیش پردازش با در حالت بدون پیش پردازش برابر با 0.7665 میشود. باتوجه به مقادیر ومقایسه آن ها میتوان نتیجه گرفت که AUC است. همچنین مقدار و در حالت پیش پردازش با خودرمنزنگار با سه لایه پنهان با AUC در حالت بدون پیش پردازش دارای بهترین مقدار logistic regression بهتر است بدون پردازش logistic regression است که برابر با 0.4953 می باشد. بنابراین الگوریتم AUC 6 و 3 نوروں دارای کمترین مقدار اجرا شود.

LDA

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5609
4 , 2	0.7205
4 , 3	0.6992
5 , 2	0.5
5 , 3	0.7099
5 , 4	0.6362
6 , 3	0.5192
6 , 4	0.7358
6 , 5	0.7205
7 , 2	0.7052
7 , 3	0.6946
7 , 4	0.7099
7 , 5	0.6959
7 , 6	0.7325

neurons	AUC
5 , 3 , 2	0.7052
5 , 4 , 2	0.5571
5 , 4 , 3	0.7052
6 , 3 , 2	0.5272
6 , 4 , 2	0.7052
6 , 5 , 2	0.5975
6 , 5 , 3	0.5272
6 , 5 , 4	0.6946
7 , 4 , 2	0.5272
7 , 4 , 3	0.5252
7 , 5 , 2	0.5013
7 , 5 , 3	0.6946
7 , 5 , 4	0.7158
7 , 6 , 2	0.5226
7 , 6 , 3	0.7205
7 , 6 , 4	0.7099
7 , 6 , 5	0.6043

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 4و نوروں است که مقداری برابر با 0.7358 دارد. در پیش پردازش خودرمنگار با سه لایه پنهان بهترین مقدار برابر با 0.7205 است که6 بهترین حالت با 6مولفه اصلی بدست می آید که مقداری برابر با PCA مربوط به تعداد 7و6و3 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 کهAUC0.7065 دارد. همچنین مقدار است بنابراین بهتر است این الگوریتم بدون پیش پردازش اجرا شود و همچنین AUCدر حالت بدون پیش پردازش دارای مقدار مطلوب تر .درحالت پیش پردازش خودرمنگار با دو لایه پنهان و تعداد 5و2 نوروں است که مقداری برابر با 0.5 داردAUCکمترین مقدار

QDA

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5106
4 , 2	0.6886
4 , 3	0.5
5 , 2	0.5
5 , 3	0.7252
5 , 4	0.5
6 , 3	0.5
6 , 4	0.5205
6 , 5	0.5
7 , 2	0.5
7 , 3	0.6992
7 , 4	0.5
7 , 5	0.7133
7 , 6	0.7252

neurons	AUC
5 , 3 , 2	0.6992
5 , 4 , 2	0.5
5 , 4 , 3	0.5
6 , 3 , 2	0.5
6 , 4 , 2	0.6992
6 , 5 , 2	0.5
6 , 5 , 3	0.5
6 , 5 , 4	0.5013
7 , 4 , 2	0.5
7 , 4 , 3	0.5
7 , 5 , 2	0.5
7 , 5 , 3	0.672
7 , 5 , 4	0.5
7 , 6 , 2	0.5
7 , 6 , 3	0.7465
7 , 6 , 4	0.5
7 , 6 , 5	0.5

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 AUC با پیش پردازش خودرمنگار با دولایه پنهان در حالت های 5 و 3 نوروں و 7 و 6 نوروں دارای بیشترین مقدار QDA در بررسی با 0.7252 است. در پیش پردازش خودرمنگار با سه لایه پنهان مقدار 0.7465 بیشترین است که مربوط 7 و 6 و 3 نوروں است. در حالت پیش آن برابر با 0.7299 است. در حالت بدون پیش پردازش AUC نیز بیشترین مقدار به حالت 3 مولفه تعلق دارد که مقدار PCA پردازش برابر با 0.7431 شده است. با توجه به مقادیر بدست آمده به این نتیجه میرسیم که پیش پردازش خودرمنگار با سه لایه پنهان و AUC مقدار برابر AUC در این حالت بهتر است و کمترین مقدار QDA مقدار 0.7465 که مربوط 7 و 6 و 3 نوروں است بهترین مقدار را دارد، بنابراین اجرای با 0.5 است که به تعدادی از حالت ها تعلق دارد

GNB

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5106
4 , 2	0.6328
4 , 3	0.666
5 , 2	0.5272
5 , 3	0.7052
5 , 4	0.6328
6 , 3	0.5047
6 , 4	0.5903
6 , 5	0.6886
7 , 2	0.6992
7 , 3	0.692
7 , 4	0.6248
7 , 5	0.6375
7 , 6	0.7405

neurons	AUC
5 , 3 , 2	0.678
5 , 4 , 2	0.5571
5 , 4 , 3	0.7039
6 , 3 , 2	0.5192
6 , 4 , 2	0.5584
6 , 5 , 2	0.5405
6 , 5 , 3	0.5371
6 , 5 , 4	0.6328
7 , 4 , 2	0.5358
7 , 4 , 3	0.5205
7 , 5 , 2	0.5119
7 , 5 , 3	0.5737
7 , 5 , 4	0.7138
7 , 6 , 2	0.5358
7 , 6 , 3	0.6401
7 , 6 , 4	0.5962
7 , 6 , 5	0.5524

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 7و6 نوروں است که مقدار آن برابر با 0.7405AUCدر حالت پیش پردازش خودرمنگار با دولایه پنهان بهترین مقدار gnb در حالت 7و5و4 بیشترین است که برابر با0.7138 می باشد AUCمی باشد. در حالت پیش پردازش خودرمنگار با سه لایه پنهان مقدار درحالت AUCدارای بیشترین مقدار برابر با 0.7265 می باشد که مربوط به 7مولفه اصلی است. همچنین مقدار PCA درحالت پیش پردازش است و AUCدر حالت بدون پیش پردازش دارای بیشترین مقدار gnb بدون پیش پردازش برابر با 0.7431 است. بنابراین میتوان گفت که مربوط به حالت پیش پردازش خودرمنگار با دو لایه پنهان و تعداد 6و3 AUCدرنتیجه دراین حالت مطلوب تر است و از طرفی کمترین مقدار نوروں برابر با 0.5047 است.

SVM

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5
4 , 2	0.7039
4 , 3	0.6992
5 , 2	0.5
5 , 3	0.7146
5 , 4	0.6149
6 , 3	0.5
6 , 4	0.7358
6 , 5	0.7252
7 , 2	0.7099
7 , 3	0.7039
7 , 4	0.6992
7 , 5	0.7112
7 , 6	0.7418

neurons	AUC
5 , 3 , 2	0.6946
5 , 4 , 2	0.5
5 , 4 , 3	0.7146
6 , 3 , 2	0.5
6 , 4 , 2	0.7031
6 , 5 , 2	0.5869
6 , 5 , 3	0.5
6 , 5 , 4	0.7099
7 , 4 , 2	0.5
7 , 4 , 3	0.5
7 , 5 , 2	0.5
7 , 5 , 3	0.6992
7 , 5 , 4	0.7205
7 , 6 , 2	0.5
7 , 6 , 3	0.7205
7 , 6 , 4	0.7099
7 , 6 , 5	0.5

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 7و6 نوروں درلایه ها است که مقدار SVM بهترین حالت برابر با 0.7205 بهترین پاسخ AUC درحالت پیش پردازش خودرمنگار با سه لایه پنهان در حالت های 7و5و4 نوروں و 7و6و3 نوروں با مقدار برابر با 0.7158 که با 6مولفه اصلی بدست می آید. در حالت بدون AUC در بهترین حالت دارای مقدار PCA در پیش پردازش SVM است. در حالت بدون پیش پردازش دارای SVM برابر با 0.7618 میشود. با توجه به اطلاعات بدست آمده میتوان گفت که AUC پیش پردازش مقدار برابر با 0.5 است که به تعدادی از حالت ها تعلق دارد AUC بهترین مقدار است که مطلوب تر است و همچنین کمترین مقدار

Random forest

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5726
4 , 2	0.6725
4 , 3	0.6466
5 , 2	0.5563
5 , 3	0.7057
5 , 4	0.6211
6 , 3	0.564
6 , 4	0.6678
6 , 5	0.6925
7 , 2	0.6292
7 , 3	0.6644
7 , 4	0.6759
7 , 5	0.6857
7 , 6	0.6785

neurons	AUC
5 , 3 , 2	0.6912
5 , 4 , 2	0.536
5 , 4 , 3	0.6611
6 , 3 , 2	0.5526
6 , 4 , 2	0.6346
6 , 5 , 2	0.5584
6 , 5 , 3	0.6045
6 , 5 , 4	0.6665
7 , 4 , 2	0.6066
7 , 4 , 3	0.5249
7 , 5 , 2	0.5466
7 , 5 , 3	0.6831
7 , 5 , 4	0.6725
7 , 6 , 2	0.5913
7 , 6 , 3	0.6818
7 , 6 , 4	0.6419
7 , 6 , 5	0.5887

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 0.7057 AUC با پیش پردازش خودرمنگار با دولایه پنهان حالت 5و3 نوروں در لایه ها با مقدار random forest بهترین حالت در روش در آن 0.6912 است. در حالت AUC است. در پیش پردازش خودرمنگار با سه لایه پنهان حالت با 5و3 و2 نوروں بهترین حالت است که مقدار برابر با 0.7418 است. درحالت 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 درنتیجه میتوان گفت اجرای الگوریتم با این روش بهتر است وکمترین مقدار نیز مربوط به PCA است که مربوط به پیش پردازش با یک مولفه است که مقدار آن 0.4735 است PCA پیش پردازش

Bagging

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5547
4 , 2	0.6857
4 , 3	0.6206
5 , 2	0.5609
5 , 3	0.7304
5 , 4	0.5679
6 , 3	0.5726
6 , 4	0.6359
6 , 5	0.6751
7 , 2	0.6598
7 , 3	0.6372
7 , 4	0.6619
7 , 5	0.6292
7 , 6	0.6878

neurons	AUC
5 , 3 , 2	0.7031
5 , 4 , 2	0.5347
5 , 4 , 3	0.6437
6 , 3 , 2	0.5679
6 , 4 , 2	0.6274
6 , 5 , 2	0.7018
6 , 5 , 3	0.5479
6 , 5 , 4	0.6253
7 , 4 , 2	0.5326
7 , 4 , 3	0.5806
7 , 5 , 2	0.5542
7 , 5 , 3	0.6266
7 , 5 , 4	0.6639
7 , 6 , 2	0.6113
7 , 6 , 3	0.7197
7 , 6 , 4	0.618
7 , 6 , 5	0.6006

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 5و3 نوروں است که مقدار آن برابرAUCدر حالت پیش پردازش خودرمزنگار با دولایه پنهان بهترین مقدار bagging در روش با 0.7304 می باشد. در پیش پردازش خودرمزنگار با سه لایه پنهان مقدار 0.7197 بیشترین است که مربوط به تعداد 7و6 نوروں است. بهترین حالت با23 مولفه اصلی بدست می آید که مقدار آن برابر 0.7179 است. در حالت بدون پیش پردازشPCA درحالت پیش پردازش با در حالت بدون پیشbagging برابر با 0.7524 بدست آمده است. با توجه به نتایج میتوان گفت که بهترین حالت برای الگوریتم AUC نیز مقدار با 1 مولفه است که مقدار آن0.4556 است PCA پردازش بدست می آید و وکمترین مقدار نیز مربوط به پیش پردازش

Decision Tree

Auto encoder pre-processing results

neurons	AUC
3 , 2	0.5453
4 , 2	0.6992
4 , 3	0.6619
5 , 2	0.529
5 , 3	0.7138
5 , 4	0.6151
6 , 3	0.5326
6 , 4	0.6904
6 , 5	0.6598
7 , 2	0.6398
7 , 3	0.7223
7 , 4	0.6593
7 , 5	0.6351
7 , 6	0.741

neurons	AUC
5 , 3 , 2	0.6891
5 , 4 , 2	0.5168
5 , 4 , 3	0.6704
6 , 3 , 2	0.5313
6 , 4 , 2	0.6359
6 , 5 , 2	0.6639
6 , 5 , 3	0.5692
6 , 5 , 4	0.6606
7 , 4 , 2	0.5619
7 , 4 , 3	0.5474
7 , 5 , 2	0.5373
7 , 5 , 3	0.6585
7 , 5 , 4	0.6665
7 , 6 , 2	0.5853
7 , 6 , 3	0.6806
7 , 6 , 4	0.6619
7 , 6 , 5	0.5853

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 AUC0.741 با پیش پردازش خودرمنگار با دولایه پنهان حالت 7و6 نوروں دارای بیشترین مقدار decision tree در بررسی است. درحالت پیش پردازش با خودرمنگار با سه لایه پنهان بهترین حالت با تعداد 5و3و2 نوروں با مقدار 0.6891 است. درحالت پیش AUC 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 است decision tree برابر با 0.7511 بدست آمده است. بنابراین میتوان گفت که بهترین حالت برای اجرا الگوریتم با 1 مولفه است که مقدار آن 0.4807 است PCA وکمترین مقدار نیز مربوط به پیش پردازش

Compared Models based on pre-processing methods

Auto encoder (with two hidden layers)

در حالت 7 و 6 نوروں دارای بیشترین SVM با بررسی نتایج مختلف در حالت پیش پردازش خودرمنگار با دولاہ پنهان متوجه میشویم که روش است که برابر با 0.7418 می باشد و همچنین مقدار 0.5 کمترین است که در چندمورد وجود دارد AUC مقدار

Auto encoder (with three hidden layers)

برابر با 0.7465 است که بزرگترین مقدار موجود در این AUC در حالت 7 و 6 و 3 نوروں دارای QDA با مشاهده نتایج بدست آمده میتوان گفت که حالت می باشد و مطلوب ترین می باشد، همچنین مقدار 0.5 کمترین است که در چندمورد وجود دارد

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 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 های

References

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

