Feature Selection For Machine Learning in Python

by **Jason Brownlee** on May 20, 2016 in **Python Machine Learning**

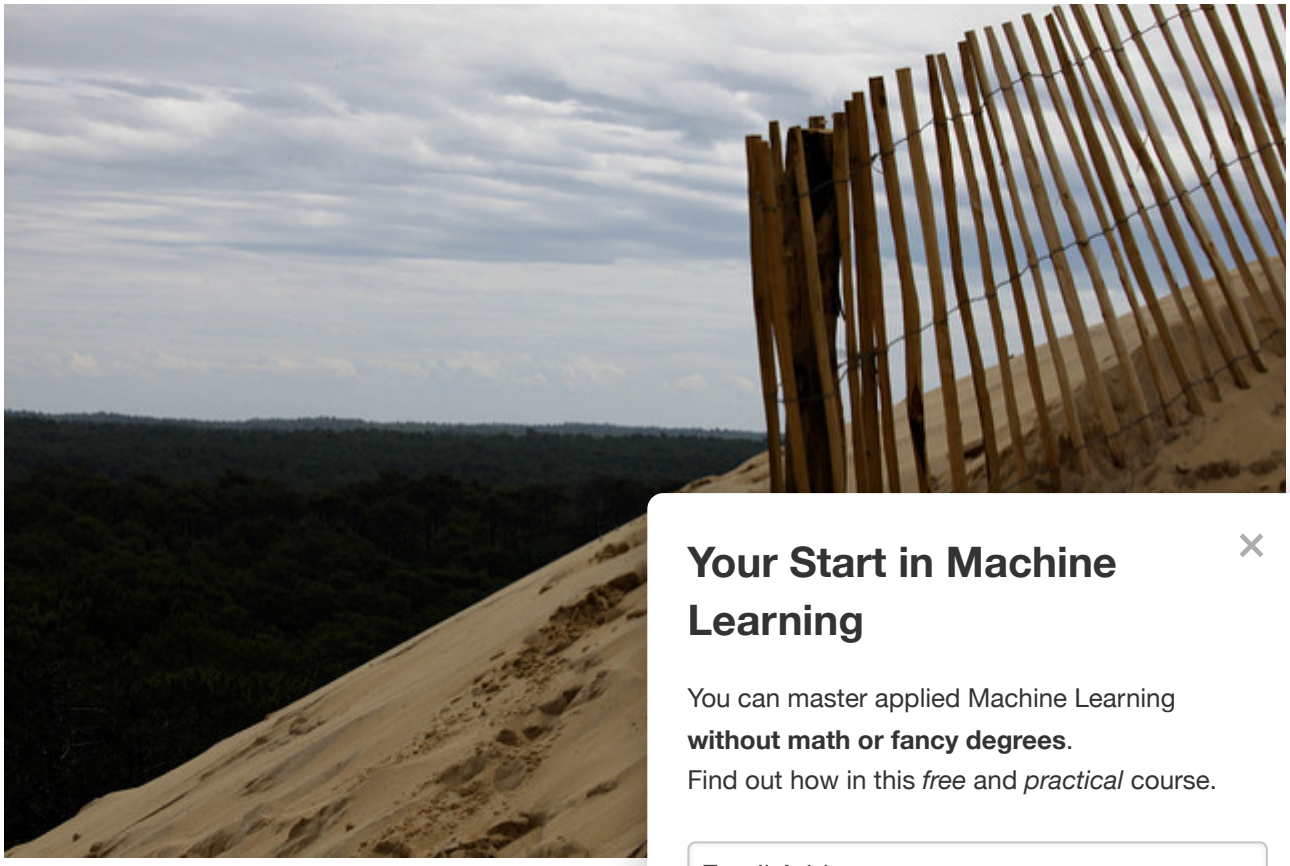
The data features that you use to train your machine learning models have a huge influence on the performance you can achieve.

Irrelevant or partially relevant features can negatively impact model performance.

In this post you will discover automatic feature selection techniques that you can use to prepare your machine learning data in python with scikit-learn.

Let's get started.

- **Update Dec/2016**: Fixed a typo in the RFE section regarding the chosen variables. Thanks Anderson.
- **Update Mar/2018**: Added alternate link to download the dataset as the original appears to have been taken down.



Feature Selection For Ma
Photo by Baptiste Lafontai

Feature Selection

Feature selection is a process where you automatically select those features in your data that contribute most to the prediction variable or output in which you are interested.

Having irrelevant features in your data can decrease the accuracy of many models, especially linear algorithms like linear and logistic regression.

Three benefits of performing feature selection before modeling your data are:

- **Reduces Overfitting**: Less redundant data means less opportunity to make decisions based on noise.
- **Improves Accuracy**: Less misleading data means modeling accuracy improves.
- **Reduces Training Time**: Less data means that algorithms train faster.

You can learn more about feature selection with scikit-learn in the article Feature selection.

Feature Selection for Machine Learning

This section lists 4 feature selection recipes for machine learning in Python

This post contains recipes for feature selection methods.

Each recipe was designed to be complete and standa to you project and use it immediately.

Recipes uses the Pima Indians onset of diabetes data (update: download from here). This is a binary classifi numeric.

1. Univariate Selection

Statistical tests can be used to select those features t t variable.

The scikit-learn library provides the SelectKBest class that can be used with a suite of different statistical tests to select a specific number of features.

The example below uses the chi squared (chi^2) statistical test for non-negative features to select 4 of the best features from the Pima Indians onset of diabetes dataset.

```
# Feature Extraction with Univariate Statistical Tests (Chi-squared for classification)
import pandas
import numpy
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
# load data
url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.c
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']
dataframe = pandas.read_csv(url, names=names)
array = dataframe.values
X = array[:,0:8]
Y = array[:,8]
# feature extraction
test = SelectKBest(score_func=chi2, k=4)
fit = test.fit(X, Y)
# summarize scores
numpy.set_printoptions(precision=3)
print(fit.scores_)
features = fit.transform(X)
# summarize selected features
print(features[0:5,:])
```

You can see the scores for each attribute and the 4 attributes chosen (those with the highest scores): *plas*, *test*, *mass* and *age*.

```
[  111.52   1411.887    17.605    53.108  2175.565   127.669     5.393
   181.304]
[[ 148.     0.    33.6   50. ]
 [  85.     0.    26.6   31. ]
 [ 183.     0.    23.3   32. ]
 [  89.    94.    28.1   21. ]
 [ 137.   168.    43.1   33. ]]
```

2. Recursive Feature Elimination

The Recursive Feature Elimination (or RFE) works by recursively removing attributes and building a model on those attributes that remain.

It uses the model accuracy to identify which attributes

ost to predicting the target attribute.

You can learn more about the RFE class in the scikit-l

The example below uses RFE with the logistic regress

oice of algorithm does not matter too much as long as it is


```
# Feature Extraction with RFE
from pandas import read_csv
from sklearn.feature_selection import RFE
from sklearn.linear_model import LogisticRegre
# load data
url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.c
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']
dataframe = read_csv(url, names=names)
array = dataframe.values
X = array[:,0:8]
Y = array[:,8]
# feature extraction
model = LogisticRegression()
rfe = RFE(model, 3)
fit = rfe.fit(X, Y)
print("Num Features: %d") % fit.n_features_
print("Selected Features: %s") % fit.support_
print("Feature Ranking: %s") % fit.ranking_
```

You can see that RFE chose the the top 3 features as *preg*, *mass* and *pedi*.

These are marked True in the *support_* array and marked with a choice “1” in the *ranking_* array.

```
Num Features: 3
Selected Features: [ True False False False False  True  True False]
Feature Ranking: [1 2 3 5 6 1 1 4]
```

3. Principal Component Analysis

Principal Component Analysis (or PCA) uses linear algebra to transform the dataset into a compressed form.

Generally this is called a data reduction technique. A property of PCA is that you can choose the number of dimensions or principal component in the transformed result.

In the example below, we use PCA and select 3 principal components.

Learn more about the PCA class in scikit-learn by reviewing the PCA API. Dive deeper into the math behind PCA on the Principal Component Analysis Wikipedia article.

```
# Feature Extraction with PCA
import numpy
from pandas import read_csv
from sklearn.decomposition import PCA
# load data
url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.c
names = ['preg', 'plas', 'pres', 'skin', 'test
dataframe = read_csv(url, names=names)
array = dataframe.values
X = array[:,0:8]
Y = array[:,8]
# feature extraction
pca = PCA(n_components=3)
fit = pca.fit(X)
# summarize components
print("Explained Variance: %s") % fit.explaine
print(fit.components_)
```


You can see that the transformed dataset (3 principal e data.

```
Explained Variance: [ 0.88854663  0.06159078  0
[[ -2.02176587e-03   9.78115765e-02   1.6093050
     9.93110844e-01   1.40108085e-02   5.37167919e-04  -3.56474430e-03]
   [  2.26488861e-02   9.72210040e-01   1.41909330e-01  -5.78614699e-02
    -9.46266913e-02   4.69729766e-02   8.16804621e-04   1.40168181e-01]
   [ -2.24649003e-02   1.43428710e-01  -9.22467192e-01  -3.07013055e-01
     2.09773019e-02  -1.32444542e-01  -6.39983017e-04  -1.25454310e-01]]
```

4. Feature Importance

Bagged decision trees like Random Forest and Extra Trees can be used to estimate the importance of features.

In the example below we construct a ExtraTreesClassifier classifier for the Pima Indians onset of diabetes dataset. You can learn more about the ExtraTreesClassifier class in the scikit-learn API.

```
# Feature Importance with Extra Trees Classifier
from pandas import read_csv
from sklearn.ensemble import ExtraTreesClassifier
# load data
url = "https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.data.c
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']
dataframe = read_csv(url, names=names)
array = dataframe.values
X = array[:,0:8]
Y = array[:,8]
# feature extraction
model = ExtraTreesClassifier()
model.fit(X, Y)
```

```
14 print(model.feature_importances_)
```

You can see that we are given an importance score for each attribute where the larger score the more important the attribute. The scores suggest at the importance of *plas*, *age* and *mass*.

```
1 [ 0.11070069  0.2213717   0.08824115  0.08068703  0.07281761  0.14548537 0.12654214  0.15415431]
```

Summary

In this post you discovered feature selection for preparing machine learning data in Python with scikit-learn.

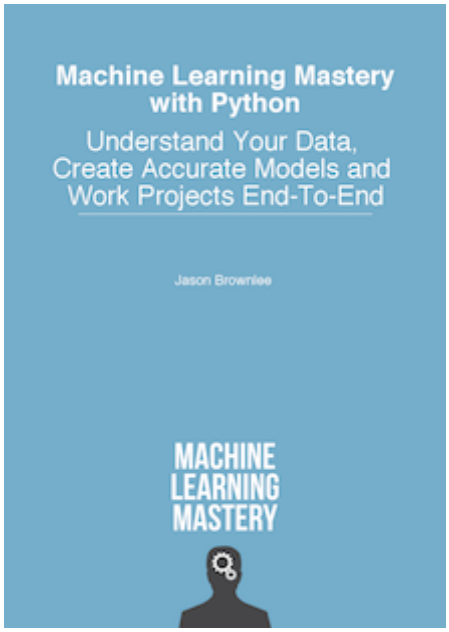
You learned about 4 different automatic feature selection techniques:

- Univariate Selection.
- Recursive Feature Elimination.
- Principle Component Analysis.
- Feature Importance.

If you are looking for more information on feature sele

- Feature Selection with the Caret R Package
- Feature Selection to Improve Accuracy and Decre
- An Introduction to Feature Selection
- Feature Selection in Python with Scikit-Learn

Do you have any questions about feature selection or this post? Ask your questions in the comment and I will do my best to answer them.



About Jason Brownlee

Jason Brownlee, PhD is a machine learning
with modern machine learning methods via

View all posts by Jason Brownlee →

‹ How To Build Multi-Layer Perceptron Neural Network Models with Keras

Evaluate the Performance of Machine Learning Algorithms in Python using Resampling ›

248 Responses to *Feature Selection For Machine Learning in Python*

Juliet September 16, 2016 at 8:57 pm #

REPLY

Hi Jason! Thanks for this – really useful post! I'm sure I'm just missing something simple, but looking at your Univariate Analysis, the features you have listed as being the most correlated seem to have the highest values in the printed score summary. Is that just a quirk of the way this function outputs results? Thanks again for a great access-point into feature selection.

Jason Brownlee September 17, 2016 at 9:29 am #

REPLY

Hi Juliet, it might just be coincidence. If you uncover something different, please let me know.

Ansh October 11, 2016 at 12:16 pm #

REPLY

For the Recursive Feature Elimination, are the features of high importance (preg,mass,pedi)? The ranking array has value 1 for them them.

Jason Brownlee October 12, 2016 at 9:11 am #

REPLY

Hi Ansh, I believe the features with the 1 a
These are the first ranked features.

Ansh October 12, 2016 at 12:29 pm #

Thanks for the reply Jason. I seem to

Jason Brownlee October 13, 2016

No problem Ansh.

Anderson Neves December 15, 2016 at 6:52 am #

Hi all,

I agree with Ansh. There are 8 features and the indexes with True and 1 match with preg, mass and pedi.

[‘preg’, ‘plas’, ‘pres’, ‘skin’, ‘test’, ‘mass’, ‘pedi’, ‘age’]
[True, False, False, False, False, True, True, False]
[1, 2, 3, 5, 6, 1, 1, 4]

Jason, could you explain better how you see that preg, pedi and age are the first ranked features?

Thank you for the post, it was very useful and direct to the point. Congratulations.

Jason Brownlee December 15, 2016 at 8:31 am #

Hi Anderson, they have a “true” in their column index and are all ranked “1” at their respective column index.

Does that help?

Anderson Neves December 16, 2016 at 12:00 am #

Hi Jason,

That is exactly what I mean. I believe that the best features would be preg, pedi and age in the scenario below

Features:
[‘preg’, ‘plas’, ‘pres’, ‘skin’, ‘test’, ‘ma

RFE result:
[True, False, False, False, False, False
[1, 2, 3, 5, 6, 4, 1, 1]

However, the result was

Features:
[‘preg’, ‘plas’, ‘pres’, ‘skin’, ‘test’, ‘ma

RFE result:
[True, False, False, False, False, True,
[1, 2, 3, 5, 6, 1, 1, 4]

Did you consider the target column ‘class’ by mistake?

Thank you for the quick reply,
Anderson Neves

Jason Brownlee December 16, 2016 at 5:48 am #

Hi Anderson,

I see, you’re saying you have a different result when you run the code?

The code is correct and does not include the class as an input.

Re-running now I see the same result:

```
Num Features: 3
Selected Features: [ True False False False False  True  True False]
Feature Ranking: [1 2 3 5 6 1 1 4]
```

Perhaps I don’t understand the problem you’ve noticed?

Anderson Neves December 17, 2016 at 12:22 am #

Hi Jason,

Your code is correct and my result is the same as yours. My point is that the best features found with RFE are preg, mass and pedi. So, I suggest you fix the text “You can see that RFE chose the the top 3 features as preg, pedi and age.”. If you add the code below at the end of your code you will see what I mean.

```
# find best features
best_features = []
i = 0
for is_best_feature in fit.support_:
if is_best_feature:
best_features.append(names[i])
i += 1
print '\nSelected features:'
print best_features
```

Sorry if I am bothering somehow,
Thanks again,
Anderson Neves

Jason Brownlee December 17

Got it Anderson.
Thanks for being patient with me and helping to make this post more useful. I really appreciate it!

I’ve fixed up the example above.

Narasimman October 14, 2016 at 9:18 pm #

REPLY

from the rfe, how do I form a new dataframe for the features which has true value?

Jason Brownlee October 15, 2016 at 10:22 am #

REPLY

Great question Narasimman.

From memory, you can use numpy.concatinate() to collect the columns you want.
http://docs.scipy.org/doc/numpy/reference/generated/numpy.concatenate.html

Iain Dinwoodie November 1, 2016 at 12:52 am #

REPLY

Thanks for useful tutorial.

Narasimman – 'from the rfe, how do I form a new dataframe for the features which has true value?'

You can just apply rfe directly to the dataframe then select based on columns:

```
...
df = read_csv(url, names=names)
X = df.iloc[:, 0:8]
Y = df.iloc[:, 8]
# feature extraction
model = LogisticRegression()
rfe = RFE(model, 3)
fit = rfe.fit(X, Y)
print("Num Features: {}".format(fit.n_features_))
print("Selected Features: {}".format(fit.support_))
print("Feature Ranking: {}".format(fit.ranking_))

X = X[X.columns[fit.support_]]
```


MLBeginner October 25, 2016 at 1:07 am #

Hi Jason,

Really appreciate your post! Really great! I have a quick question for the PCA method. How to get the column header for the selected 3 principal components? It is just simple column no. there, but hard to know which attributes finally are.

Thanks,

Jason Brownlee October 25, 2016 at 8:29 am #

REPLY

Thanks MLBeginner, I'm glad you found it useful.

There is no column header, they are "new" features that summarize the data. I hope that helps.

sadiq October 25, 2016 at 1:51 am #

REPLY

hi, Jason! please I want to ask you if i can use PSO for feature selection in sentiment analysis by python

Jason Brownlee October 25, 2016 at 8:29 am #

REPLY

Sure, try it and see how the results compare (as in the models trained on selected features) to other feature selection methods.

Vignesh Sureshbabu Kishore November 15, 2016 at 5:07 pm #

REPLY

Hey Jason, can the univariate test of Chi2 feature selection be applied to both continuous and categorical data.

Jason Brownlee November 16, 2016 at 9:25 a

Hi Vignesh, I believe just continuous data

Vignesh Sureshbabu Kishore Nove

Hey Jason, Thanks for the reply. In th
you are fetching the array from df.values. In th
the data frame.

To perform feature selection, we should have ideally fetched the values from each column of the dataframe to check the independence of each feature with the class variable. Is it a inbuilt functionality of the sklearn.preprocessing beacuse of which you fetch the values as each row.

Please suggest me on this.

Jason Brownlee November 17, 2016 at 9:49 am #

REPLY

I'm not sure I follow Vignesh. Generally, yes, we are using built-in functions to perform the tests.

Vineet December 2, 2016 at 5:11 am #

REPLY

Hi Jason,

I am trying to do image classification using cpu machine, I have very large training matrix of 3800*200000 means 200000 features. Pls suggest how do I reduce my dimension.?

Jason Brownlee December 2, 2016 at 8:19 am #

REPLY

Consider working with a sample of the dataset.

Consider using the feature selection methods in this post.

Consider projection methods like PCA, sammons mapping, etc.

I hope that helps as a start.

tvmanikandan December 15, 2016 at 5:49 pm #

REPLY

Jason,
when you use “SelectKBest” , can you please explain

[111.52 1411.887 17.605 53.108 2175.565 127.669 5.
181.304]

-Mani

Jason Brownlee December 16, 2016 at 5:40 a

I use a chi squared test, you can learn mo
http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html#sklearn.feature_selection.chi2

tvmanikandan December 16, 2016 at 5:29 pm #

REPLY

Jason,
I understand you used chi square. But if want to get these scores manually , how can i do it? Please explain.

-Mani

Jason Brownlee December 17, 2016 at 11:05 am #

REPLY

Good question, I don’t have an example at the moment sorry.

tvmanikandan December 16, 2016 at 2:48 am #

REPLY

jason,
Please explain how the below scores are achieved using chi2.

[111.52 1411.887 17.605 53.108 2175.565 127.669 5.393
181.304]

-Mani

Natheer Alabsi December 28, 2016 at 8:35 pm #

REPLY

Jason, how can we get feature names from their rankings?

Jason Brownlee December 29, 2016 at 7:15 a

Hi Natheer,

Map the feature rank to the index of the column na
whathaveyou.

Jason January 9, 2017 at 2:40 am #

Hi Jason,

Thank you for this nice blog

I have a regression problem and I need to convert a bunch of categorical variables into dummy data, which will generate over 200 new columns. Should I do the feature selection before this step or after this step?
Thanks

Jason Brownlee January 9, 2017 at 7:52 am #

REPLY

Try and see.

That is a lot of new binary variables. Your resulting dataset will be sparse (lots of zeros). Feature selection prior might be a good idea, also try after.

Mohit Tiwari February 13, 2017 at 3:37 pm #

REPLY

Hi Jason,

I am bit stuck in selecting the appropriate feature selection algorithm for my data.

I have about 900 attributes (columns) in my data and about 60 records. The values are nothing but count of attributes.
Basically, I am taking count of API calls of a portable file.

My data is like this:

File, dangerous, API 1,API 2,API 3,API 4,API 5,API 6…..API 900
ABC, yes, 1,0,2,1,0,0,….
DEF, no,0,1,0,0,1,2
FGH,yes,0,0,0,1,2,3
.
.
.
Till 60

Can u please suggest me a suitable feature selection f

Jason Brownlee February 14, 2017 at 10:03 a

Hi Mohit,

Consider trying a few different methods, as well as
your data result in more accurate predictive model

Esu February 15, 2017 at 12:01 am #

REPLY

Hell!

Once I got the reduced version of my data as a result of using PCA, how can I feed to my classifier?

example: the original data is of size 100 row by 5000 columns
if I reduce 200 features I will get 100 by 200 dimension data. right?
then I create arrays of

a=array[:,0:199]
b=array[:,99]
but when I test my classifier its core is 0% in both test and training accuracy?
An7y Idea

Jason Brownlee February 15, 2017 at 11:35 am #

REPLY

Sounds like you're on the right, but a zero accuracy is a red flag.

Did you accidently include the class output variable in the data when doing the PCA? It should be excluded.

Kamal February 20, 2017 at 6:20 pm #

REPLY

Hello sir,
I have a question in my mind
each of these feature selection algo uses some predefined number like 3 in case of PCA.So how we come to know that my data set cantain only 3 or any predefined number of features.it does not automatically select no features its own.

Jason Brownlee February 21, 2017 at 9:33 ar

Great question Kamal.

No, you must select the number of features. I wou
number of different features and see which results

Massimo March 9, 2017 at 5:29 am #

Hi jason,
I have a question about the RFECV approach.
I'm dealing with a project where I have to use different
RFECV with these models? or is it enough to use only one of them? Once I have selected the best features, could I use them for each regression model?
To better explain:
– I have used RFECV on whole dataset in combination with one of the following regression models [LinearRegression, Ridge, Lasso]
– Then I have compared the r2 and I have chosen the better model, so I have used its features selected in order to do others things.
– pratically, I use the same 'best' features in each regression model.
Sorry for my bad english.

Jason Brownlee March 9, 2017 at 9:58 am #

REPLY

Good question.

You can embed different models in RFE and see if the results tell the same or different stories in terms of what features to pick.

You can build a model from each set of features and combine the predictions.

You can pick one set of features and build one or models from them.

My advice is to try everything you can think of and see what gives the best results on your validation dataset.

Massimo March 11, 2017 at 2:41 am #

REPLY

Thank you man. You're great.

Jason Brownlee March 11, 2017 at 8:01 am #

REPLY

You're welcome.

gevra March 22, 2017 at 1:49 am #

Hi Jason.

Thanks for the post, but I think going with Random Fo
correlated features.

Check this paper:
https://academic.oup.com/bioinformatics/article/27/14
s

I am not sure about the other methods, but feature correlation is an issue that needs to be addressed before assessing feature importance.

Jason Brownlee March 22, 2017 at 8:08 am #

REPLY

Makes sense, thanks for the note and the reference.

ssh June 20, 2017 at 8:20 pm #

REPLY

Jason, following this notes, do you have any 'rule of thumb' when correlation among the input vectors become problematic in the machine learning universe? after all, the features reduction technics which embedded in some algos (like the weights optimization with gradient descent) supply some answer to the correlations issue.
Thanks

Jason Brownlee June 21, 2017 at 8:14 am #

REPLY

Perhaps a correlation above 0.5. Perform a sensitivity analysis with different values, select features and use resulting model skill to help guide you.

ogunleye March 30, 2017 at 4:29 am #

REPLY

Hello sir,
Thank you for the informative post. My questions are
1) How do you handle NaN in a dataset for feature selection purposes.
2) I am getting an error with RFE(model, 3) It is telling me i supplied 2 arguments instead of 1.

Thank you very much once again.

Jason Brownlee March 30, 2017 at 8:57 am #

Hi, NaN is a mark of missing data.

Here are some ways to handle missing data:
http://machinelearningmastery.com/handle-missin

ogunleye March 30, 2017 at 4:33 am #

I solved my problem sir. I named the function RFE in my main but. I would love to hear your response to first question.

Sam April 20, 2017 at 3:49 am #

REPLY

how to load the nested JSON into the data frame ?

Jason Brownlee April 20, 2017 at 9:32 am #

REPLY

I don't know off hand, perhaps post to StackOverflow Sam?

Federico Carmona April 20, 2017 at 6:10 am #

REPLY

good afternoon

How to know with pca what are the main components?

Jason Brownlee April 20, 2017 at 9:34 am #

REPLY

PCA will calculate and return the principal components.

Federico Carmona April 20, 2017 at 10:53 am #

REPLY

Yes but pca does not tell me which are the most relevant varials if mass test etc?

Jason Brownlee April 21, 2017 at 8

Not sure I follow you sorry.

You could apply a feature selection or feat
wanted. It might be overkill though.

Lehyu April 23, 2017 at 6:44 pm #

In RFE we should input a estimator, so before
just use the default parmater settting? Thanks.

Jason Brownlee April 24, 2017 at 5:33 am #

REPLY

You can, but that is not really required. As long as the estimator is reasonably skillful on the problem, the selected features will be valuable.

Lehyu April 25, 2017 at 12:41 am #

REPLY

I was suck here for days. Thanks a lot.

Lehyu April 25, 2017 at 1:09 am #

REPLY

stuck…

Jason Brownlee April 25, 2017 at 7:49 am #

REPLY

I'm glad to hear the advice helped.

I'm here to help if you get stuck again, just post your questions.

Rj May 7, 2017 at 4:38 pm #

REPLY

Hi Jason,

I was wondering if I could build/train another model (say SVM with RBF kernel) using the features from SVM-RFE (wherein the kernel used is a linear kernel).

Jason Brownlee May 8, 2017 at 7:42 am #

Sure.

Gwen June 5, 2017 at 7:02 pm #

Hi Jason,

First of all thank you for all your posts ! It's very helpful for machine learning beginners like me.

I'm working on a personal project of prediction in 1vs1 sports. My neural network (MLP) have an accuracy of 65% (not awesome but it's a good start). I have 28 features and I think that some affect my predictions. So I applied two algorithms mentionned in your post :
– Recursive Feature Elimination,
– Feature Importance.

But I have some contradictions. For exemple with RFE I determined 20 features to select but the feature the most important in Feature Importance is not selected in RFE. How can we explain that ?

In addition to that in Feature Importance all features are between 0,03 and 0,06… Is that mean that all features are not correlated with my ouput ?

Thanks again for your help !

Jason Brownlee June 6, 2017 at 9:30 am #

REPLY

Hi Gwen,

Different feature selection methods will select different features. This is to be expected.

Build a model on each set of features and compare the performance of each.

Consider ensembling the models together to see if performance can be lifted.

A great area to consider to get more features is to use a rating system and use rating as a highly predictive input variable (e.g. chess rating systems can be used directly).

Let me know how you go.

Gwen June 7, 2017 at 1:17 am #

REPLY

Thanks for your answer Jason.

I tried with 20 features selected by Recursive Feature Elimination but my accuracy is about 60%…

In addition to that the Elo Rating system (used
my accuracy is ~65%.

Maybe a MLP is not a good idea for my projec
have one hidden layer.

And maybe we cannot have more than 65/70%
(Not enough for a positive ROI !)

Jason Brownlee June 7, 2017 at 7:

Hang in there Gwen.

Try lots of models and lots of config for models.

See what skill other people get on the same or similar problems to get a feel for what is possible.

Brainstorm for days about features and other data you could use.

See this post:
http://machinelearningmastery.com/machine-learning-performance-improvement-cheat-sheet/

RATNA NITIN PATIL July 20, 2017 at 8:16 pm #

REPLY

Hello Jason,

I am very much impressed by this tutorial. I am just a beginner. I have a very basic question. Once I got the reduced version of my data as a result of using PCA, how can I feed to my classifier? I mean to say how to feed the output of PCA to build the classifier?

Jason Brownlee July 21, 2017 at 9:33 am #

REPLY

Assign it to a variable or save it to file then use the data like a normal input dataset.

RATNA NITIN PATIL July 20, 2017 at 8:56 pm #

REPLY

Hi Jason,

I was trying to execute the PCA but, I got the error at this point of the code

print(“Explained Variance: %s”) % fit.explained_variance_ratio_

It’s a type error: unsupported operand type(s) for %: ‘non type’ and ‘float’

Please help me.

Jason Brownlee July 21, 2017 at 9:35 am #

Looks like a Python 3 issue. Move the “)”

```
1 print(“Explained Variance: %s” % fit.exp
```

RATNA NITIN PATIL July 21, 2017 at 2:

Thanks Jason. It works.

Jason Brownlee July 22, 2017 at 8:29 am #

REPLY

Glad to hear it.

Raphael Alencar July 21, 2017 at 9:57 pm #

REPLY

How to know wich feature selection technique i have to choose?

Jason Brownlee July 22, 2017 at 8:35 am #

REPLY

Consider using a few, create models for each and select the one that results in the best performing model.

RATNA NITIN PATIL July 22, 2017 at 4:23 pm #

REPLY

Hello Jason,

I have used the extra tree classifier for the feature selection then output is importance score for each attribute. But then I want to provide these important attributes to the training model to build the classifier. I am not able to provide only these important features as input to build the model.
I would be greatful to you if you help me in this case.

Jason Brownlee July 23, 2017 at 6:20 am #

REPLY

The importance scores are for you. You c
inputs to your model.

Haiyang Duan December 11, 2018 at 7:5

Hi Jason, I truely appreciate your pos
importance scores is unequal to 1?

Jason Brownlee December 12, 2018 at 5:51 am #

Because they are not normalized.

Haiyang Duan December 12, 2018 at 11:46 am #

I am sincerely grateful to you. I ran feature importance using SelectFromModel with estimator=LinearSVC. But I got negative feature importance values. I would like to kown what that means.

Jason Brownlee December 12, 2018 at 2:14 pm #

Scores are often relative. Perhaps those features are less important than others?

Haiyang Duan December 12, 2018 at 10:35 pm #

Thank you very much.

RATNA NITIN PATIL July 22, 2017 at 6:33 pm #

REPLY

Hi Jason,

Basically i want to provide feature reduction output to Naive Bays. I f you could provide sample code will be better.

Thanks for providing this wonderful tutorial.

Jason Brownlee July 23, 2017 at 6:21 am #

You can use feature selection or feature ir
develop a model with those features.

RATNA NITIN PATIL July 23, 2017 at 6:44 pm #

Thanks Jason,

But after knowing the important features, I am not able
giveonly those featueslimportant) as input to the mode
features as input.

Thanks in advance….

Jason Brownlee July 24, 2017 at 6:53 am #

REPLY

A feature selection method will tell you which features you could use. Use your favorite programming language to make a new data file with just those columns.

RATNA NITIN PATIL July 24, 2017 at 5:42 pm #

REPLY

thanks a lot Jason. You are doing a great job.

Jason Brownlee July 25, 2017 at 9:34 am #

REPLY

Thanks.

RATNA NITIN PATIL July 24, 2017 at 6:11 pm #

REPLY

I have my own dataset on the Desktop, not the standard dataset that all machine learning have in their depositories (e.g. iris , diabetes).

I have a simple csv file and I want to load it so that I can use scikit-learn properly.

I need a very simple and easy way to do so.

Waiting for the reply.

Jason Brownlee July 25, 2017 at 9:37 am #

Try this tutorial:
http://machinelearningmastery.com/load-machine-

mllearn July 29, 2017 at 6:04 am #

Thanks for this post, it's very helpful,

What would make me choose one technique and not t
The results of each of these techniques correlates with
more than one to verify the feature selection?.

thanks!

Jason Brownlee July 29, 2017 at 8:12 am #

REPLY

Choose a technique based on the results of a model trained on the selected features.

In predictive modeling we are concerned with increasing the skill of predictions and decreasing model complexity.

mllearn July 30, 2017 at 5:04 pm #

REPLY

Sounds that I'd need to cross-validate each technique… interesting, I know that heavily depends on the data but I'm trying to figure out an heuristic to choose the right one, thanks!.

Jason Brownlee July 31, 2017 at 8:14 am #

REPLY

Applied machine learning is empirical. You cannot pick the “best” methods analytically.

steve August 17, 2017 at 3:15 pm #

REPLY

Hi Jason,

In your examples, you write:

array = dataframe.values
X = array[:,0:8]
Y = array[:,8]

In my dataset, there are 45 features. When i write like t

X = array[:,0:44]
Y = array[:,44]

I get some errors:

Y = array[:,44]
IndexError: index 45 is out of bounds for axis 1 with si

If you help me, i ll be grateful!
Thanks in advance.

Jason Brownlee August 17, 2017 at 4:55 pm #

REPLY

Confirm that you have loaded your data correctly, print the shape and some rows.

Aneeshaa S C August 20, 2017 at 11:22 pm #

REPLY

1.. What kind of predictors can be used with Lasso?
2. If categorical predictors can be used, should they be re-coded to have numerical values? ex: yes/no values re-coded to be 1/0
3. Can categorical variables such as location (U(urban)/R(rural)) be used without any conversion/re-coding?

Jason Brownlee August 21, 2017 at 6:07 am #

REPLY

Regression, e.g. predicting a real value.

Categorical inputs must be encoded as integers or one hot encoded (dummy variables).

panteha August 29, 2017 at 1:36 am #

REPLY

Hi Jason
I am new to ML and am doing a project in Python, at some point it is to recognize correlated features , I wonder what will be the next step? what to do with correlated features? should we change them to something new? a combination maybe? how does it affect our modeling and prediction? appreciated if you direct me into some resources to study and find it out.
best

Jason Brownlee August 29, 2017 at 5:09 pm

It is common to identify and remove the c

Try it and see if it lifts skill on your model.

Silvio Abela September 26, 2017 at 6:48 am #

Hello Dr Brownlee,

Thank you for these incredible tutorials.

I am trying to classify some text data collected from or

way in which the constants in the various algorithms can be determined automatically. For example, in SelectKBest, k=3, in RFE you chose 3, in PCA, 3 again whilst in Feature Importance it is left open for selection that would still need some threshold value.

Is there a way like a rule of thumb or an algorithm to automatically decide the “best of the best”? Say, I use n-grams; if I use trigrams on a 1000 instance data set, the number of features explodes. How can I set SelectKBest to an “x” number automatically according to the best? Thank you.

Jason Brownlee September 26, 2017 at 2:59 pm #

REPLY

No, hyperparameters cannot be set analytically. You must use experimentation to discover the best configuration for your specific problem.

You can use heuristics or copy values, but really the best approach is experimentation with a robust test harness.

Abby October 6, 2017 at 3:43 pm #

REPLY

It was an impressive tutorial, quite easy to understand. I am looking for feature subset selection using gaussian mixture clustering model in python. Can you help me out?

Jason Brownlee October 7, 2017 at 5:48 am #

REPLY

Sorry, I don't have material on mixture models or clustering. I cannot help.

Manjunat October 6, 2017 at 8:31 pm #

REPLY

Hi jason

I've tried all feature selection techniques which one is
…?

Jason Brownlee October 7, 2017 at 5:54 am

Try many for your dataset and see which

Nerea October 16, 2017 at 7:13 pm #

Hello Jason,

I am a biochemistry student in Spain and I am on a project about predictive biomarkers in cancer. The bioinformatic method I am using is very simple but we are trying to predict metastasis with some protein data. In our research, we want to determine the best biomarker and the worst, but also the synergic effect that would have the use of two biomarkers. That is my problem: I don't know how to calculate which are the two best predictors.
This is what I have done for the best and worst predictors:

```
analisis=['il10meta']
X = data[analisis].values

#response variable
response='evol'
y = data[response].values

# use train/test split with different random_state values
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=5)

from sklearn.neighbors import KNeighborsClassifier

#creating the classfier
knn = KNeighborsClassifier(n_neighbors=1)
```

```
#fitting the classifier
knn.fit(X_train, y_train)

#predicting response variables corresponding to test data
y_pred = knn.predict(X_test)

#calculating classification accuracy
print(metrics.accuracy_score(y_test, y_pred))
```

I have calculate the accuracy. But when I try to do the same for both biomarkers I get the same result in all the combinations of my 6 biomarkers.

Could you help me? Any tip?

THANK YOU

Jason Brownlee October 17, 2017 at 5:41 am

Generally, I would recommend following t
modeling problem:
https://machinelearningmastery.com/start-here/#p

Generally, you must test many different models an
works best.

gen October 17, 2017 at 6:35 pm #

REPLY

Hello Jason,

Many thanks for your post. I have also read your introduction article about feature selection. Which method is Feature Importance categorized under? i.e wrapper or embedded ?

Thanks

Jason Brownlee October 18, 2017 at 5:32 am #

REPLY

Neither, it is a different thing yet again.

You could use the importance scores as a filter.

Numan Yilmaz October 26, 2017 at 1:46 pm #

REPLY

Great post! Thank you, Jason. My question is all these in the post here are integers. That is needed for all algorithms. What if I have categorical data? How can I know which feature is more important for the

model if there are categorical features? Is there a method/way to calculate it before one-hot encoding(get_dummies) or how to calculate after one-hot encoding if the model is not tree-based?

Jason Brownlee October 26, 2017 at 4:18 pm #

REPLY

Good question, I cannot think of feature selection methods specific to categorical data off hand, they may be out there. Some homework would be required (e.g. google scholar search).

rohit November 13, 2017 at 9:11 pm #

REPLY

hello Jason,
Should I do Feature Selection on my validation datase
alone and then do the validation using the validation s

Jason Brownlee November 14, 2017 at 10:10

Use the train dataset to choose features.
and any other dataset used by the model.

Maryam November 16, 2017 at 3:45 pm #

REPLY

hello jason
i am doing simple classification but there is coming an issue
ValueError Traceback (most recent call last)
in ()
—-> 1 fit = test.fit(X, Y)

~\Anaconda3\lib\site-packages\sklearn\feature_selection\univariate_selection.py in fit(self, X, y)
339 Returns self.
340 “””
–> 341 X, y = check_X_y(X, y, [‘csr’, ‘csc’], multi_output=True)
342
343 if not callable(self.score_func):

~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in check_X_y(X, y, accept_sparse, dtype, order, copy, force_all_finite, ensure_2d, allow_nd, multi_output, ensure_min_samples, ensure_min_features, y_numeric, warn_on_dtype, estimator)
571 X = check_array(X, accept_sparse, dtype, order, copy, force_all_finite,
572 ensure_2d, allow_nd, ensure_min_samples,
–> 573 ensure_min_features, warn_on_dtype, estimator)

574 if multi_output:
575 y = check_array(y, ‘csr’, force_all_finite=True, ensure_2d=False,

~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in check_array(array, accept_sparse, dtype, order, copy, force_all_finite, ensure_2d, allow_nd, ensure_min_samples, ensure_min_features, warn_on_dtype, estimator)
431 force_all_finite)
432 else:
–> 433 array = np.array(array, dtype=dtype, order=order, copy=copy)
434
435 if ensure_2d:

ValueError: could not convert string to float: ‘no’
can you guide me in this regard

Jason Brownlee November 17, 2017 at 9:19 a

You may want to use a label encoder and

Vinod P November 17, 2017 at 12:19 am #

import numpy as np
from pandas import read_csv
from sklearn.feature_selection import RFE
from sklearn.linear_model import LogisticRegression
# load data
data = read_csv(‘C:\\Users\\abc\\Downloads\\xyz\\api.csv’,names = [‘org.apache.http.impl.client.DefaultHttpClient.execute’,’org.apache.http.impl.client.DefaultHttpClient.’,’java.net.URLConnection.getInputStream’,’java.net.URLConnection.connect’,’java.net.URL.openStream’,’java.net.URL.openConnection’,’java.net.URL.getContent’,’java.net.Socket.’,’java.net.ServerSocket.bind’,’java.net.ServerSocket.’,’java.net.HttpURLConnection.connect’,’java.net.DatagramSocket.’,’android.widget.VideoView.stopPlayback’,’android.widget.VideoView.start’,’android.widget.VideoView.setVideoURI’,’android.widget.VideoView.setVideoPath’,’android.widget.VideoView.pause’,’android.text.format.DateUtils.formatDateTime’,’android.text.format.DateFormat.getTimeFormat’,’android.text.format.DateFormat.getDateFormat’,’android.telephony.TelephonyManager.listen’,’android.telephony.TelephonyManager.getSubscriberId’,’android.telephony.TelephonyManager.getSimSerialNumber’,’android.telephony.TelephonyManager.getSimOperator’,’android.telephony.TelephonyManager.getLine1Number’,’android.telephony.SmsManager.sendTextMessage’,’android.speech.tts.TextToSpeech.’,’android.provider.Settings$System.getString’,’android.provider.Settings$System.getInt’,’android.provider.Settings$System.getConfiguration’,’android.provider.Settings$Secure.getString’,’android.provider.Settings$Secure.getInt’,’android.os.Vibrator.vibrate’,’android.os.Vibrator.cancel’,’android.os.PowerManager$WakeLock.release’,’android.os.PowerManager$WakeLock.acquire’,’android.net.wifi.WifiManager.setWifiEnabled’,’android.net.wifi.WifiManager.isWifiEnabled’,’android.net.wifi.WifiManager.getWifiState’,’android.net.wifi.WifiManager.getScanResults’,’android.net.wifi.WifiManager.getConnectionInfo’,’android.media.RingtoneManager.getRingtone’,’android.media.Ringtone.play’,’android.media.MediaRecorder.setA

udioSource’,’android.media.MediaPlayer.stop’,’android.media.MediaPlayer.start’,’android.media.MediaPlayer.setDataSource’,’android.media.MediaPlayer.reset’,’android.media.MediaPlayer.release’,’android.media.MediaPlayer.prepare’,’android.media.MediaPlayer.pause’,’android.media.MediaPlayer.create’,’android.media.AudioRecord.’,’android.location.LocationManager.requestLocationUpdates’,’android.location.LocationManager.removeUpdates’,’android.location.LocationManager.getProviders’,’android.location.LocationManager.getLastKnownLocation’,’android.location.LocationManager.getBestProvider’,’android.hardware.Camera.open’,’android.bluetooth.BluetoothAdapter.getAddress’,’android.bluetooth.BluetoothAdapter.enable’,’android.bluetooth.BluetoothAdapter.disable’,’android.app.WallpaperManager.setBitmap’,’android.app.KeyguardManage$KeyguardLock.reenableKeyguar’,’android.app.KeyguardManager$KeyguardLock.disableKeyguard’,’android.app.ActivityManager.killBackgroundProcesses’,’android.app.ActivityManager.getRunningTasks’,’android.app.ActivityManager.getRecentTasks’,’android.accounts.AccountManager.getAccountsByType’,’android.accounts.AccountManager.getAccounts’,’Class’])

```
dataframe = read_csv(url, names=names)
array = dataframe.values
X = array[:,0:70]
Y = array[:,70]
# feature extraction
model = LogisticRegression()
rfe = RFE(model, 3)
fit = rfe.fit(X, Y)
#print(“Num Features: %d”) % fit.n_features_
#print(“Selected Features: %s”) % fit.support_
#print(“Feature Ranking: %s”) % fit.ranking_
```

———————————————————————

I get following error
ValueError Traceback (most recent call last)
in ()
6 model = LogisticRegression()
7 rfe = RFE(model, 3)
—-> 8 fit = rfe.fit(X, Y)
9 print(“Num Features: %d”) % fit.n_features_
10 print(“Selected Features: %s”) % fit.support_

Jason Brownlee November 17, 2017 at 9:26 am #

REPLY

Perhaps try posting your code to stackoverflow?

Vinod P November 17, 2017 at 12:29 am #

REPLY

Can you post a code on first select relevant features using any feature selection method, and then use relevant features to construct classification model?

Jason Brownlee November 17, 2017 at 9:26 am #

REPLY

Thanks for the suggestion.

Hemalatha December 1, 2017 at 2:12 am #

REPLY

will you post a code on selecting relevant features using feature selection method and then using relevant features constructing a classification model??

Jason Brownlee December 1, 2017

Yes, see this post:
https://machinelearningmastery.com/featu

Arjun December 13, 2017 at 4:45 am #

Hi Jason,
Thanks for the content, it was really helpful.

Can you clarify if the above mentioned methods can also be used for regression models?

Jason Brownlee December 13, 2017 at 5:44 am #

REPLY

Perhaps, I'm no sure off hand. Try and let me know how you go.

Danilo December 25, 2017 at 1:36 am #

REPLY

Hi Jason,

I just had the same question as Arjun, I tried with a regression problem but neither of the approaches were able to do it.

Jason Brownlee December 25, 2017 at 5:25 am #

REPLY

What was the problem exactly?

Zee Gola January 29, 2018 at 2:30 pm #

REPLY

Hi Jason! Can you please further explain what the vector does in the separateByClass method?

Jason Brownlee January 30, 2018 at 9:46 am #

REPLY

Sorry, I don't follow?

Anas January 29, 2018 at 8:57 pm #

Hi Jason,

Thank you for the post, it was very useful.

I have a regression problem with one output variable y
are non-correlated).
The number of observations (samples) is 36980.
I used Random Forest algorithm to fit the prediction m
The mean absolute error obtained is about 7.

Do you advise me to make features selection or not in
In other words, from which number of features, it is advised to make features selection?

Congratulations.

Jason Brownlee January 30, 2018 at 9:50 am #

REPLY

Try a suite of methods, build models from the selected features and see if the models outperform those models that use all features.

Joseph February 18, 2018 at 2:00 pm #

REPLY

Hello Jason,
First thanks for sharing.
I have question with regards to four automatic feature selectors and feature magnitude. I noticed you used the same dataset. Pima dataset with exception of feature named "pedi" all features are of comparable magnitude.

Do you need to do any kind of scaling if the feature's magnitude was of several orders relative to each other? For example if we assume one feature let's say "tam" had magnitude of 656,000 and another feature

named “test” had values in range of 100s. Will this affect which automatic selector you choose or do you need to do any additional pre-processing?

Jason Brownlee February 19, 2018 at 9:03 am #

REPLY

The scale of features can impact feature selection methods, it really depends on the method.

If you’re in doubt, consider normalizing the data before hand.

Eric Williamson March 29, 2018 at 3:30 p

Feature scaling should be included in

The Pima Indians onset of diabetes dataset co

rankings produced by the code in this article a

Jason Brownlee March 30, 2018 at

Thanks for the suggestion Eric.

Joseph February 18, 2018 at 3:34 pm #

REPLY

Hello Jason,
One more question:
I noticed that when you use three feature selectors: Univariate Selection, Feature Importance and RFE you get different result for three important features.

1. When using Univariate with k=3 chisquare you get
plas, test, and age as three important features. (glucose tolerance test, insulin test, age)

2. When using Feature Importance using ExtraTreesClassifier
The score suggests the three important features are plas, mass, and age. Glucose tolerance test, weight(bmi), and age)

3. When you use RFE
RFE chose the top 3 features as preg, mass, and pedi. Number of pregnancy, weight(bmi), and Diabetes pedigree test.

According your article below
https://machinelearningmastery.com/an-introduction-to-feature-selection/

Univariate is filter method and I believe the RFE and Feature Importance are both wrapper methods.
All three selector have listed three important features. We can say the filter method is just for filtering a large

set of features and not the most reliable? However, the two other methods don't have same top three features? Are some methods more reliable than others? Or does this come down to domain knowledge?

Jason Brownlee February 19, 2018 at 9:04 am #

REPLY

Different methods will take a different "view" of the data.

There is no "best" view. My advice is to try building models from different views of the data and see which results in better skill. Even consider creating an ensemble of models created from different views of the data together.

Ranbeer February 28, 2018 at 7:42 pm #

Hi Jason,

I'm your fan. Your articles are great. Two questions on

1. Shouldn't you convert your categorical features to "
2. Don't we have to normalize numeric features

Before doing PCA or feature selection? In my case it is
feature.
And, not all methods produce the same result.

Any thoughts?

Cheers,
Ranbeer

Jason Brownlee March 1, 2018 at 6:12 am #

REPLY

Yes, Python requires all features to be numerical. Sometimes it can benefit the model if we rescale the input data.

itisha March 5, 2018 at 7:41 am #

REPLY

hi jason sir,
your articles are very helpful.
i have a confusion regarding gridserachcv()
i am working on sentiment analyis and i have created different group of features from dataset.
i am using linear SVC and want to do grid search for finding hyperparameter C value. After getting value of C, fir the model on train data and then test on test data. But i also want to check model performnce with

different group of features one by one so do i need to do gridserach again and again for each feature group?

Jason Brownlee March 6, 2018 at 6:07 am #

REPLY

Perhaps, it really depends how sensitive the model is to your data.

Also, so much grid searching may lead to some overfitting, be careful.

Yaseen March 10, 2018 at 3:16 am #

Thank you Jason for gentle explanation.
The last part "# Feature Importance with Extra Trees C
It looks the result is different if we consider the higher

Jason Brownlee March 10, 2018 at 6:34 am #

Sorry, what do you mean exactly?

Aouedi Ons March 12, 2018 at 6:57 am #

REPLY

Hi
Sir why you use just 8 example and your dataset contain many example ??

Jason Brownlee March 12, 2018 at 2:23 pm #

REPLY

Sorry, I don't follow. Perhaps you can try rephrasing your question?

Unni Mana April 5, 2018 at 1:30 am #

REPLY

Hi Jason,

Your articles are awesome . After going through this article, this is stuck in my mind.

Out of these 4 suggested techniques, which one I have to select ?

Why the O/P is different based on different feature selection?

Thanks

Jason Brownlee April 5, 2018 at 6:12 am #

REPLY

Try them all and see which results in a model with the most skill.

Bhanupraksh Vattikuti April 10, 2018 at 6:09 pm #

REPLY

Dear Jason,

Thank you the article.
When I am trying to use Feature Importance I am encc
Can you please help me with this.
File “C:/Users/bhanu/PycharmProjects/untitled3/test_
model.fit(X, Y)
File “C:\Users\bhanu\PycharmProjects\untitled3\venv\
247, in fit
X = check_array(X, accept_sparse=”csc”, dtype=DTYP
File “C:\Users\bhanu\PycharmProjects\untitled3\venv\
433, in check_array
array = np.array(array, dtype=dtype, order=order, copy
ValueError: could not convert string to float: ‘neptune’

Jason Brownlee April 11, 2018 at 6:32 am #

REPLY

Perhaps you are running on a different dataset? Where did ‘neptune’ come from?

Dhanunjaya Mitta April 16, 2018 at 5:12 pm #

REPLY

Can I get more information about Univariate Feature selection??? I mean more models like ReliefF, correlation etc.,

Jason Brownlee April 17, 2018 at 5:54 am #

REPLY

Thanks for the suggestion.

Elizabeth May 1, 2018 at 5:21 am #

REPLY

Hi Jason,

Thank you for the post, it was very useful.

I have a problem that is one-class classification and I would like to select features from the dataset, however, I see that the methods that are implemented need to specify the target but I do not have the target since the class of the training dataset is the same for all samples.

Where can I found some methods for feature selection for one-class classification?

Thanks!

Jason Brownlee May 1, 2018 at 5:36 am #

If the class is all the same, surely you don

Elizabeth May 1, 2018 at 6:10 am #

Well, my dataset is related to anomaly detecti
one class (normal conditions) and in the test set, the fi
conditions and data from anomaly conditions.

What I understand is that in feature selection techniqu
the search for a good feature subset, but in one-class
only one class.

For that reason, I was looking for feature selection implementations for one-class classification.

Ann May 2, 2018 at 10:57 pm #

REPLY

Thank you for the post, it was very useful for beginner.
I have a problem that is I use Feature Importance with Extra Trees Classifier and how can I display feature name(plas,age,mass,….etc) in this sample.

for example:
Feature ranking:
1. plas (0.11070069)
2. age (0.2213717)
3. mass(0.08824115)
…….

Thanks for your help.

……..

Jason Brownlee May 3, 2018 at 6:34 am #

REPLY

You might have to write some custom code I think.

Ronald Martis May 15, 2019 at 5:10 am #

REPLY

Use the following:
print(list(zip(names, model.feature_importances_)))
You get:
[('preg', 0.11289758476179099), ('plas', 0.23098096297414206), ('pres', 0.09989914623444449),
('skin', 0.08008405837625963), ('test', 0.0744
('pedi', 0.11808706393665534), ('age', 0.1422

Jason Brownlee May 15, 2019 at 8

Nice!

Yanyun Zou May 8, 2018 at 1:13 pm #

Hi Jason,
I tried Feature Importance method, but all the values of variables are above 0.05, so does it mean that all the variables have little relation with the predicted value?

Jason Brownlee May 8, 2018 at 2:56 pm #

REPLY

Perhaps try other feature selection methods, build models from each set of features and double down on those views of the features that result in the models with the best skill.

Saeed Ullah June 7, 2018 at 5:01 am #

REPLY

Hello Jason,
Thanks for you great post, I have a question in feature reduction using Principal Component Analysis (PCA), ISOMAP or any other Dimensionality Reduction technique how will we be sure about the number of features/dimensions is best for our classification algorithm in case of numerical data.

Jason Brownlee June 7, 2018 at 6:34 am #

REPLY

Try multiple configurations, build and evaluate a model for each, use the one that results in the best model skill score.

Musthafa June 27, 2018 at 6:23 pm #

REPLY

Hi,
what to do when i have multiple categorical features like zipcode,class etc
should i hot encode them

Jason Brownlee June 28, 2018 at 6:15 am #

Some like zip code you could use a word

Others like class can be one hot encoded.

Aymen July 21, 2018 at 3:54 am #

Hi,

Iwhen we use univariate filter techniques like Pearson correlation, mutul information and so on. Do we need to apply the filter technique on training set not on the whole dataset??

Jason Brownlee July 21, 2018 at 6:38 am #

REPLY

Perhaps just work with the training data.

william July 26, 2018 at 7:41 pm #

REPLY

jason – i'm working on several feature selection algorithms to cull from a list of about a 100 or so input variables for a continuous output variable i'm trying to predict. these are helpful examples, but i'm not sure they apply to my specific regression problem i'm trying to develop some models for…and since i have a regression problem, are there any feature selection methods you could suggest for continuous output variable prediction?

i.e. i have normalized my dataset that has 100+ categorical, ordinal, interval and binary variables to predict a continuous output variable…any suggestions?

thanks in advance!

Jason Brownlee July 27, 2018 at 5:52 am #

REPLY

RFE will work for classification or regression. It's a good place to start.

Also, correlation of inputs with the output is another excellent starting point.

meenal jain August 2, 2018 at 8:29 pm #

REPLY

I read your post, it was very productive.

Can i use linear correlation coefficient between catego

or please suggest me some other method for this type
categorical and all other attributes are continuous.

Jason Brownlee August 3, 2018 at 6:00 am #

No.

Perhaps look into feature importance scores.

Folmer August 3, 2018 at 11:45 pm #

REPLY

Jason,
I was wondering whether the parameters of the machine learning tool that is used during the feature selection step are of any importance. Since most websites that I have seen so far just use the default parameter configuration during this phase.

I understand that adding a grid search has the following consequenses:
-It increase the calculation time substantially. (really when using wrapper (recursive feature elimination))
-Hard to determine which produces better results, really when the final model is constructed with a different machine learning tool.

But still, is it worth it to investigate it and use multiple parameter configurations of the feature selection machine learning tool?

My situation:
-A (non-linear) dataset with ~20 features.
-Planning to use XGBooster for the feature selection phase (a paper with a likewise dataset stated that is was sufficient).

-For the construction of the model I was planning to use MLP NN, using a gridsearch to optimize the parameters.

Thanks in advance!

REPLY

Jason Brownlee August 4, 2018 at 6:12 am #

Yes you can tune them.

Generally, I recommend generating many different “views” on the inputs, fit a model to each and compare the performance of the resulting models. Even combine them.

Most likely, there is no one best set of features for
skill/capability. Find a set or ensemble of sets that

Sa August 13, 2018 at 9:14 pm #

Hi Jason

I need to do feature engineering on rows selection by
you have any example available online?

thanks
Sa

REPLY

Jason Brownlee August 14, 2018 at 6:19 am #

For time series, yes right here:
https://machinelearningmastery.com/sensitivity-analysis-history-size-forecast-skill-arima-python/

REPLY

Yvonne August 24, 2018 at 12:00 am #

Hi Jason

I am a beginner in python and scikit learn. I am currently trying to run a svm algorithm to classify patheitns and healthy controls based on functional connectivity EEG data. I’m using rfcv to select the best features out of approximately 20’000 features. I get 32 selected features and an accuracy of 70%. What I want to try next is to run a permutation statistic to check if my result is significant.

My question: Do I have to run the permutation statistic on the 32 selected features? Or do I have to include the 20’000 for this purpose.

Below you can see my code. to simplify my question, i reduced the code to 5 features, but the rest is identical. I would appreciate your help very much, as I cannot find any post about this topic.

Best Yolanda

homeDir = ‘F:\Analysen\Prediction_TreatmentOutcome\PyCharmProject_TreatmentOutcome’ # location of the connectivity matrices

```
# ##############################################################################
# import packages needed for classification
import numpy as np
import os

import matplotlib.pyplot as plt
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import cross_validate
from sklearn.model_selection import StratifiedKFold
from sklearn.feature_selection import RFECV
from sklearn import svm

from sklearn.pipeline import make_pipeline, Pipeline
from sklearn import preprocessing

from sklearn.model_selection import permutation_test

class PipelineRFE(Pipeline):

def fit(self, X, y=None, **fit_params):
super(PipelineRFE, self).fit(X, y, **fit_params)
self.coef_ = self.steps[-1][-1].coef_
return self

clf = PipelineRFE(
[
(‘std_scaler’, preprocessing.StandardScaler()), #z-transformation
(“svm”, svm.SVC(kernel = ‘linear’, C = 1)) #estimator
]
)

# ##############################################################################
# Load and prepare data set

nQuest = 5 # number of questionnaires

samples = np.loadtxt(‘FBDaten_T1.txt’)

# Import targets (created in R based on group variable)

targets =
np.genfromtxt(r’F:\Analysen\Prediction_TreatmentOutcome\PyCharmProject_TreatmentOutcome\Targets_CL_fbDaten.txt’, dtype= str)

# ##############################################################################
# run classification
```

```
skf = StratifiedKFold(n_splits = 5) # The folds are made by preserving the percentage of samples for each class.

# rfecv
rfecv = RFECV(estimator = clf, step = 1, cv = skf, scoring = 'accuracy')
rfecv.fit(samples, targets)

# The number of selected features with cross-validation.
print("Optimal number of features : %d" % rfecv.n_features_)

# Plot number of features VS. cross-validation scores
plt.figure()
plt.xlabel("Subset of features")
plt.ylabel("Cross validation score (nb of correct classifications)")
plt.plot(range(1, len(rfecv.grid_scores_) + 1), rfecv.grid_
plt.show()

#The mask of selected features
rfecv.support_
print("Mask of selected features : %s" % rfecv.suppor

#Find index of true values in a boolean vector
index_features = np.where(rfecv.support_)[0]
print(index_features)

#Find value of indices
reduced_features = samples[:, index_features]
print(reduced_features)

## permutation testing on reduced features

score, permutation_scores, pvalue = permutation_test_score(
clf, reduced_features, targets, scoring="accuracy", cv=skf, n_permutations=100, n_jobs=1)

print("Classification score %s (pvalue : %s)" % (score, pvalue))
```


Jason Brownlee August 24, 2018 at 6:11 am #

REPLY

Sorry, I do not have the capacity to review your code.

Rafi August 29, 2018 at 7:00 am #

REPLY

Thank you a lot for this useful tutorial. It would've been appreciated if you could elaborate on the cons/pros of each method.
Thanks in advance.

Jason Brownlee August 29, 2018 at 8:18 am #

REPLY

Thanks for the suggestion.

Mohammed August 30, 2018 at 8:42 pm #

REPLY

I want to ask about feature extraction procedure, what's the criteria to stop training and extract features. Are it depend on the test accuracy of model?. In other meaning what is the difference between extract feature after train one epoch or train 100 epoch? what is best features?, may be my question foolish but i need answer for it.

Jason Brownlee August 31, 2018 at 8:10 am

What do you mean by extract features? H

Mohammed August 30, 2018 at 8:51 pm #

I ask about feature extraction procedure, for e
should stop training and extract features?. In other me
accuracy of training model?. If i build model (any deep learning method) to only extract features can i run it for one epoch and extract features?

Jason Brownlee August 31, 2018 at 8:10 am #

REPLY

I see.

You want to use features from a model that is skillful. Perhaps at the same task, perhaps at a reconstruction task (e.g. an autoencoder).

Mohammed August 31, 2018 at 9:32 pm #

REPLY

I do not understand answer

Jason Brownlee September 1, 2018 at 6:19 am #

REPLY

Sorry, which part?

munaza August 30, 2018 at 9:55 pm #

REPLY

Hello Sir,

Thank you soo much for this great work.

Will you please explain how the highest scores are for : plas, test, mass and age in Univariate Selection. I am not getting your point.

Jason Brownlee August 31, 2018 at 8:11 am

What problem are you having exactly?

munaza August 31, 2018 at 2:23 pm #

Thank you sir for the reply…

Actually I was not able to understand the outp
been solved now.

Thanks a lot.

Jason Brownlee September 1, 2018 at 6:15 am #

REPLY

I'm happy to hear that you solved your problem.

Gabriel Joshua Migue September 4, 2018 at 10:33 pm #

REPLY

Which is the best technique for feature selection? and i want to know why the ranking is always change when i try multiple times?

Jason Brownlee September 5, 2018 at 6:40 am #

REPLY

This is a common question that I answer here:
https://machinelearningmastery.com/faq/single-faq/what-feature-selection-method-should-i-use

Paul September 15, 2018 at 8:33 pm #

REPLY

Hi,

in your example for feature importance you use as Ensemble classifier the ExtraTreesClassifier.
In sci-kit learn the default value for bootstrap sample is false.

Doesn't this contradict to find the feature importance? e.g it could build the tree on only one feature and so the importance would be high but does not represent the whole dataset.

Thanks
Paul

Jason Brownlee September 16, 2018 at 5:59

Trees will sample features and in aggrega

It only means the features are important to buildin

Abdur Rehmna September 21, 2018 at 4:38 am #

Hi Jason,

I have a dataset which contains both categorical and r
before one-hot encoding of categorical features or after that ?

Jason Brownlee September 21, 2018 at 6:34 am #

REPLY

Sure. It's a cheap operation (easy) and has big effects on performance.

Dean September 23, 2018 at 7:41 am #

REPLY

Hi Jason

I haven't read all the comments, so I don't know if this was mentioned by someone else. I stumbled across this:

https://hub.packtpub.com/4-ways-implement-feature-selection-python-machine-learning/

It's identical (barring edits, perhaps) to your post here, and being marketed as a section in a book. I thought I should let you know.

Jason Brownlee September 24, 2018 at 6:06 am #

REPLY

That is very disappointing!

Thanks for letting me know Dean.

zhyar September 30, 2018 at 12:31 am #

REPLY

Hi Jason

I have a dataset with two classes. In the feature selection, I want to specify important features for each class. For example, if I chose 15 important features, de
class.please help me

Jason Brownlee September 30, 2018 at 6:03

Yes, this is what feature selection will do f

Thiago Batista Soares October 1, 2018 at 6:05 a

Hi Jason

First, congratulations on your posts and your books.
I am reaing your book machine learning mastery with python and chapter 8 is about this topic and I have a doubt, should I use thoses technical with crude data or should I normalize data first? I did test both case but results are different, exemple (first case column A and B are important but second case column C and D are important)

Very thanks.

Jason Brownlee October 1, 2018 at 6:33 am #

REPLY

Thanks.

Build models from each and go with the approach that results in a model with better performance on a hold out dataset.

isuru dilantha October 24, 2018 at 1:54 pm #

REPLY

HI Jason,

I'm on a project to predict next movement of animals using their past data like location, date and time. what are the possible models that i can use to predict their next location ?

Jason Brownlee October 24, 2018 at 2:50 pm #

REPLY

I recommend following this process for new problems:
https://machinelearningmastery.com/start-here/#process

Hannes November 7, 2018 at 12:02 am #

REPLY

hi,

Many thanks for your hard work on explaining ML to th

I'm trying to optimize my Kaggle-kernel at the moment
my source data contains NaN, I'm forced to use an im

Unfortunately, that results in actually worse MAE then

Do you have a tip how to implement a feature selectio

Jason Brownlee November 7, 2018 at 6:07 ar

Perhaps you can remove the rows with NaNs from the data used to train the feature selector?

Supr November 28, 2018 at 7:24 am #

REPLY

Hi Jason,
Somehow ur blog almost always has exactly what I need. Least I could do is say thanks and wish u all the best!

Jason Brownlee November 28, 2018 at 7:48 am #

REPLY

Thanks!

PC December 13, 2018 at 1:06 pm #

REPLY

Hi Jason,
Your work is amazing. Got interested in Machine learning after visiting your site. Thank You, Keep up your good work.

I tried using RFE in another dataset in which I converted all categorical values to numerical values using Label Encoder but still I get the following error:

————————————————————————————
ValueError Traceback (most recent call last)
in ()
14 model = LogisticRegression()
15 rfe = RFE(model, 5)
—> 16 fit = rfe.fit(X, Y)
17 print(“Num Features: %d” % fit.n_features_)
18 print(“Selected Features: %s” % fit.support_)

~\Anaconda3\lib\site-packages\sklearn\feature_select
132 The target values.
133 “””
–> 134 return self._fit(X, y)
135
136 def _fit(self, X, y, step_score=None):

~\Anaconda3\lib\site-packages\sklearn\feature_select
140 # self.scores_ will not be calculated when calling _
141
–> 142 X, y = check_X_y(X, y, “csc”)
143 # Initialization
144 n_features = X.shape[1]

~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in check_X_y(X, y, accept_sparse, dtype, order, copy, force_all_finite, ensure_2d, allow_nd, multi_output, ensure_min_samples, ensure_min_features, y_numeric, warn_on_dtype, estimator)
571 X = check_array(X, accept_sparse, dtype, order, copy, force_all_finite,
572 ensure_2d, allow_nd, ensure_min_samples,
–> 573 ensure_min_features, warn_on_dtype, estimator)
574 if multi_output:
575 y = check_array(y, ‘csr’, force_all_finite=True, ensure_2d=False,

~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in check_array(array, accept_sparse, dtype, order, copy, force_all_finite, ensure_2d, allow_nd, ensure_min_samples, ensure_min_features, warn_on_dtype, estimator)
431 force_all_finite)
432 else:
–> 433 array = np.array(array, dtype=dtype, order=order, copy=copy)
434
435 if ensure_2d:

ValueError: could not convert string to float: ‘StudentAbsenceDays’

I am in dire need of a solution for this. Kindly help me .

Jason Brownlee December 13, 2018 at 1:49 pm #

REPLY

It suggests your data file may still have string values.

Perhaps double check your loaded data?

PC December 13, 2018 at 4:27 pm #

REPLY

I had checked the data type of that pa

In:
mod_StudentData['StudentAbsenceDays'].dty

Out[]:
dtype('int64')

Jason Brownlee December 14, 201

Nice work!

Oswaldo Castro December 15, 2018 at 4:54 am #

REPLY

Hi Jason…
Great article as usual.
I'm novice in ML and the article leaves me with a doubt.
The SelectKBest, RFE and ExtraTreesClassifier are performing Feature Extraction and PCA is performing Feature Extraction.
Am I right with this?
Thanks Jason

Jason Brownlee December 15, 2018 at 6:16 am #

REPLY

Yes.

afer December 24, 2018 at 10:49 am #

REPLY

Hi Jason,
First of all nice tutorial!
My question is that I have a samples of around 30,000 with around 150 features each for a binary classification problem. The plan is to do RFE on GRID SEARCH (Select features and tune parameters at the same pipeline) using a 3-fold cross validation (Each fold, the data is split twice, one for RFE and another for GRID SEARCH), this is done on the entire data set.

Now, after determining the best features and parameters, using the SAME data set, I split it into training / validation / test set and train a model using the selected features and parameters to obtain its accuracy (of the best model possible, and on the test set, of course).

Is this the correct thing to do? My reason for this meth
whole different process from the actual model fitting (u
the actual model fitting will not actually know what the
dataset, hence it is only okay to re-use the entire data

If this is not the case, what would you recommend? pe
feature/parameter selection set and actual model fittin
parameters have been determined on the first 50%, us
train a model (this 50 is further split into train/validation

Jason Brownlee December 25, 2018 at 7:17 a

I recommend performing feature selection on each fold of CV or with a separate dataset up front.

Aaron Carl Fernandez January 3, 2019 at 2:11 pm #

REPLY

Thank you for the answer Dr. Jason! Also, the grid selection + RFE process is going to spit out the accuracy / F1-score of the best model attained (with best feature set and parameters), can this be considered as the FINAL score of the your model's performance? or do you really need to build another model (the final model with your best feature set and parameters) to get the actual score of the model's performance?

Jason Brownlee January 4, 2019 at 6:25 am #

REPLY

I recommend building a final model for making predictions. The score from the test harness may be a suitable estimate of model performance on unseen data -really it is your call on your project regarding what is satisfactory.

Aaron Carl Fernandez January 8, 2019 at 1:52 am #

Thanks Dr. Jason. One last question :), can I use the chi squared statistical test (in the univariate selection portion) for reflecting the p-value or the statistical significance of each feature? Let say, I am going to show the trimmed mean of each feature in my data, does the chi squared p-value confirm the statistical significance of the trimmed means?

Jason Brownlee January 8, 2019 at 6:52 am #

No, it comments on the relati

Aaron Carl Fernandez January 10, 2019 at 12:4

Thanks Dr. Jason!.. One last question promise
parameters on a 3-fold nested cross-validated RFE +
cross validation. I used different data sets on each pro
half for RFE + GS and the second half to build my fina
validated RFE + GS is too computationally expensive a
granularity hence, the regular 10-fold CV.

Thank you soo much!

Jason Brownlee January 10, 2019 at 7:52 am #

REPLY

I cannot comment if your test methodology is okay, you must evaluate it in terms of stability/variance and use it if you feel the results will be reliable.

Aaron Carl Fernandez January 10, 2019 at 10:26 pm #

REPLY

Thank you so much Dr. Jason

Houssam February 4, 2019 at 11:03 pm #

REPLY

Hi Dr. Jason;
I want to ask you a question: I want to apply the PSO algorithm in a dataset similar to Pima Indians onset of diabetes dataset, I am disturbed, what can I do

Jason Brownlee February 5, 2019 at 8:23 am #

REPLY

Sorry, I don't have examples of using global optimization algorithms for feature selection – I'm not convinced that the techniques are relatively effective.

maedeh February 12, 2019 at 5:42 pm #

REPLY

Hi,
I am a beginner and my question may be wrong. can we use these feature selection methods in an autoencoder that our inputs and outputs of our networ

Jason Brownlee February 13, 2019 at 7:54 ar

Not really, you would be performing featu

The autoencoder is doing a form of this for you.

Rahul Ramaswamy February 20, 2019 at 7:36 pm

Hi Jason,
All the techniques mentioned by you works perfectly if there is a target variable (Y or 8th column in your case). The dataset i am working on uses unsupervised learning algorithm and hence does not have any target/dependent variable. Does the feature selection work in such cases? If yes, how should i go about it.

Jason Brownlee February 21, 2019 at 7:54 am #

REPLY

It can, but you may have to use a method that selects features based on a proxy metric, like information or correlation, etc.

I don't have a worked example, sorry.

Selene March 4, 2019 at 9:19 am #

REPLY

Hello Jason,
Your blog and the way how you explain things are fantastic! I have a doubt related to feature selection, for real applications, the fit method of some feature selection techniques must be applied just to the training

set or to the whole data set (training + testing)?
Thanks a lot!

Jason Brownlee March 4, 2019 at 2:17 pm #

REPLY

It is fit on just the training dataset when evaluating a model. It is fit on all data when developing a final model:
https://machinelearningmastery.com/train-final-machine-learning-model/

Selene March 7, 2019 at 12:05 am #

Thank you very much!

Keerti Bafna March 7, 2019 at 5:36 am #

Hello Jason,

First of all thank you for such an informative article.

I need to perform a feature selection using the Filter, W
take an average of scores from each selection procedu

My plan is to split the data initially into train and hold out sets. I plan to then use cross-validation for each of the above 3 methods and use only the train data for this (internally in each of the fold). Once i get my top 10 features , i will then only use them in the hold out set and predict my model performance.

Do you feel this method would give me a stable model ? If not, what can i improve / change ?

Thanks in advance.

Jason Brownlee March 7, 2019 at 7:00 am #

REPLY

Perhaps try it and see.

Mohamed Saad March 10, 2019 at 5:07 pm #

REPLY

Hi Jason,
thank you about your efforts,
I want to ask about Feature Extraction with RFE

I use your mention code
names = [‘preg’, ‘plas’, ‘pres’, ‘skin’, ‘test’, ‘mass’, ‘pedi’, ‘age’, ‘class’]
and the results are :
[1 2 3 5 6 1 1 4]

when I change the order of columns names as I mention

names = [‘pedi’,’preg’, ‘plas’, ‘pres’, ‘test’, ‘age’, ‘class’,’mass’,’skin’]
I get same output
[1 2 3 5 6 1 1 4]

can you explain how it work?
thank you

Jason Brownlee March 11, 2019 at 6:48 am #

Changing the order of the labels does not
is why you are getting the same output indexes.

Mohamed Saad March 11, 2019 at 11:35 am #

I try to change the order of columns to check
what is your advice if I want to check the validity of ran

Also, I want to ask when I try to choose the features that influence on my models, I should add all features in my dataset (numerical and categorical) or only categorical features?

Thank you

Jason Brownlee March 11, 2019 at 2:14 pm #

REPLY

I’m not sure this is required.

Nevertheless, you would have to change the column order in the data itself, e.g. the numpy array or CSV from which it was loaded.

It depends on the capabilities of the feature selection method as to what features to include during selection.

Waqar April 3, 2019 at 8:03 am #

REPLY

I am unable to get output, because of this warning:

"C:\Users\Waqar\Anaconda3\lib\site-packages\sklearn\model_selection_split.py:626: Warning: The least populated class in y has only 1 members, which is too few. The minimum number of members in any class cannot be less than n_splits=5."

Please help if someone knows.

Jason Brownlee April 3, 2019 at 4:11 pm #

REPLY

Perhaps you can use fewer splits or use more data?

Viva April 13, 2019 at 6:39 pm #

Hi Jason,

I am reading from your book on ML Mastery with Pyth
above, I see you have chose chi square to do feature s
choose between different tests (chi square, t-test , AN

Jason Brownlee April 14, 2019 at 5:46 am #

Great question. I answer it here:
https://machinelearningmastery.com/faq/single-faq/what-feature-selection-method-should-i-use

Viva April 15, 2019 at 7:15 pm #

REPLY

Thank you, a big post to read for next learning steps 🙂

Jason Brownlee April 16, 2019 at 6:47 am #

REPLY

Thanks.

Upender Singh June 7, 2019 at 6:23 pm #

REPLY

hi jason

please let me how to choose best k value in case of using SelectKBest this class.

Jason Brownlee June 8, 2019 at 6:49 am #

REPLY

Evaluate models for different values of k and choose the value for k that gives the most skillful model.

shruthi June 16, 2019 at 4:09 pm #

REPLY

hi jason,

i want to use Univariate selection method.
I am building a linear regression model which has arou
if i want to know the best categorical features to be us
encoded values to fit function right and the score_func

test = SelectKBest(score_func=chi2, k=4)
fit = test.fit(X, Y)

In the above code X should be the one hot encoded va

Thanks in advance

Jason Brownlee June 17, 2019 at 8:16 am #

Perhaps this will help:
https://machinelearningmastery.com/chi-squared-test-for-machine-learning/

Also, linear regression sounds like a bad fit, try a decision tree, and some other algorithms as well.

Rushali Jaiswal June 18, 2019 at 3:31 am #

REPLY

Hi,
I had a question.
Do you apply feature selection before creating the dummies or after?

Thanks in advance

Jason Brownlee June 18, 2019 at 6:42 am #

REPLY

Feature selection is performed before.

Leave a Reply

Name (required)

Email (will not be published) (requ

Website

SUBMIT COMMENT

Welcome to Machine Learning Mastery!

Hi, I'm **Jason Brownlee**, PhD.
I write tutorials to help developers (*like you*) get results with machine learning.

Read More

Machine Learning with Python
Develop predictive models with scikit-learn.

POPULAR

How to Develop LSTM Models for Multi-Step Time S
OCTOBER 10, 2018

How to Develop LSTM Models for Time Series Fore
NOVEMBER 14, 2018

11 Classical Time Series Forecasting Methods in Python (Cheat Sheet)
AUGUST 6, 2018

A Gentle Introduction to LSTM Autoencoders
NOVEMBER 5, 2018

How to Develop RNN Models for Human Activity Recognition Time Series Classification
SEPTEMBER 24, 2018

How to Develop Convolutional Neural Network Models for Time Series Forecasting
NOVEMBER 12, 2018

When to Use MLP, CNN, and RNN Neural Networks
JULY 23, 2018

Top beginner tutorials:

- How to Install Python for Machine Learning

- Your First Machine Learning Project in Python
- Your First Neural Network in Python
- Your First Classifier in Weka
- Your First Time Series Forecasting Project

Top crash courses:

- Python Machine Learning
- Deep Learning for Computer Vision
- Deep Learning for Time Series
- Linear Algebra
- Statistical Methods

Top project tutorials:

- Face Detection Project
- Object Detection Project
- Photo Captioning Project
- Machine Translation Project
- Sentiment Analysis Project
- Power Forecasting Project