Classification

Prerequisites:

It is assumed that you will install and take a look at the following packages in python before heading to activities:

- sklearn (http://scikit-learn.org/stable/)
- (http://flask.pocoo.org/) pandas (https://pandas.pydata.org/)

This lab makes use of the iris dataset (https://archive.ics.uci.edu/ml/datasets/iris) . This dataset has four features including sepal_length, sepal_width, petal_length, and petal_length of three species of flowers: setosa, versicolor, and virginica.

Activity-1:

Description : Train and test a KNN classifier

Steps :

1. Load the iris dataset
2. Split the dataset into test and train datasets; 70% of the data should be used for training the classifier and the rest for testing
3. Train a KNN (http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html) classifier by fitting on the train dataset
4. Based on the trained classifier, predict the classes of the test dataset;
5. Print the predictions and the real class labels

 (https://github.com/mysilver/COMP9321-Data-Services/blob/master/Week9_Classification/activity_2.py)

Activity-2:

Description : Evaluate KNN classifier over iris dataset

Steps :

1. Make a copy of the code your created in the previous activity, and after prediction of the testing dataset, do as follow
2. Calculate and print the accuracy of the classifier (http://scikit-learn.org/stable/modules/generated/sklearn.metrics.accuracy_score.html)
3. Print the confusion_matrix (http://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html)

4. Calculate and print the mean precision (http://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_score.html#sklearn.metrics.precision_score) , and recall (http://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html)

(https://github.com/mysilver/COMP9321-Data-Services/blob/master/Week9_Classification/activity_2.py)

Activity-3:

Description : Evaluate various classifiers using k-fold cross evaluation

Steps :

1. Load the iris dataset
2. Perform a 5-fold cross validation (https://en.wikipedia.org/wiki/Cross-validation_(statistics)) over the following classifiers, and record their accuracy:

```
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score
from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.utils import shuffle
KNeighborsClassifier()
DecisionTreeClassifier()
LinearDiscriminantAnalysis()
LogisticRegression()
GaussianNB()
SVC()
```

3. Sort the classifiers based on their accuracy
4. Print the classifiers and their accuracy in order.

(https://github.com/mysilver/COMP9321-Data-Services/blob/master/Week9_Classification/activity_3.py)

Resource created about a month ago (Monday 14 March 2022, 03:02:59 PM), last modified 23 days ago (Monday 04 April 2022, 11:38:25 AM).

Comments

(/COMP9321/22T1/forums/search?forum_choice=resource/74090)

(/COMP9321/22T1/forums/resource/74090)

Add a comment

Solomon Rachamim (/users/z5375417) 18 days ago (Fri Apr 08 2022 16:16:11 GMT+0800 (China Standard Time)), last modified 18 days ago (Fri Apr 08 2022 16:31:27 GMT+0800 (China Standard Time))

For those who are on mac M1 I had an issue downloading sklearn (Mac M1).

I followed this and it helped:

https://stackoverflow.com/questions/68620927/insta... (https://stackoverflow.com/questions/68620927/installing-scipy-and-scikit-learn-on-apple-m1)

happy learning!

Reply

Zheng Fu (/users/z5285691) 11 days ago (Fri Apr 15 2022 18:51:00 GMT+0800 (China Standard Time))

does conda works?

Reply