Supplement (rfcv.py, rfcvgini.py, evalcross.py, rfcross.py)

==================rfcv.py==================

# -\*- coding: utf-8 -\*-

import sys,codecs

import numpy as np

import pandas as pd

import sys

filename='24\_27rand.csv'

trees=493

taco=pd.read\_csv(filename)

taco.fillna(0,inplace=True)

crossv=10

print('filename:',filename)

print("data instances and parameters:",taco.shape)

print("cross-validation:",crossv)

print("trees:",trees)

y = taco['Normal']

X=taco.drop(['Normal','Event Id'],axis=1)

from sklearn.ensemble import RandomForestClassifier

from sklearn.model\_selection import cross\_val\_score

from sklearn.metrics import \*

from sklearn.model\_selection import ShuffleSplit

cv = ShuffleSplit(n\_splits=crossv, test\_size=0.09, random\_state=54)

clf=RandomForestClassifier(criterion='entropy',n\_estimators=trees, max\_depth=None,min\_samples\_split=2, random\_state=54,n\_jobs=-1)

scores = cross\_val\_score(clf, X, y, cv=cv)

print(scores,scores.mean(),round(scores.std(),6))

===================rfcvgini.py======================

# -\*- coding: utf-8 -\*-

import sys,codecs

import numpy as np

import pandas as pd

import sys

filename='24\_27rand.csv'

trees=493

taco=pd.read\_csv(filename)

taco.fillna(0,inplace=True)

crossv=10

print('filename:',filename)

print("data instances and parameters:",taco.shape)

print("cross-validation:",crossv)

print("trees:",trees)

y = taco['Normal']

X=taco.drop(['Normal','Event Id'],axis=1)

from sklearn.ensemble import RandomForestClassifier

from sklearn.model\_selection import cross\_val\_score

from sklearn.metrics import \*

from sklearn.model\_selection import ShuffleSplit

cv = ShuffleSplit(n\_splits=crossv, test\_size=0.09, random\_state=54)

clf=RandomForestClassifier(criterion='gini',n\_estimators=trees, max\_depth=None,min\_samples\_split=2, random\_state=54,n\_jobs=-1)

scores = cross\_val\_score(clf, X, y, cv=cv)

print(scores,scores.mean(),round(scores.std(),7))

======================evalcross.py========================

import random

import subprocess as sp

for i in range(100):

j=random.randrange(150,500)

print(j)

command="python rfcross.py 24\_27rand.csv "+str(j)

sp.call(command,shell=True)

======================rfcross.py==========================

# -\*- coding: utf-8 -\*-

import sys,codecs

import numpy as np

import pandas as pd

import sys

filename=sys.argv[1]

trees=int(sys.argv[2])

taco=pd.read\_csv(filename)

taco.fillna(0,inplace=True)

crossv=10

print('filename:',filename)

print("data instances and parameters:",taco.shape)

print("cross-validation:",crossv)

print("trees:",trees)

y = taco['Normal']

X=taco.drop(['Normal','Event Id'],axis=1)

from sklearn.ensemble import RandomForestClassifier

from sklearn.model\_selection import cross\_val\_score

from sklearn.metrics import \*

from sklearn.model\_selection import ShuffleSplit

cv = ShuffleSplit(n\_splits=crossv, test\_size=0.09, random\_state=54)

clf=RandomForestClassifier(criterion='entropy',n\_estimators=trees, max\_depth=None,min\_samples\_split=2, random\_state=54,n\_jobs=-1)

scores = cross\_val\_score(clf, X, y, cv=cv)

print(scores,scores.mean(),scores.std())