Comparaison de LOF, OC-SVM et Isolation Forest

Jeu de données

Nous utilisons le jeu de données KDD99CUP

Métriques de comparaison

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from time import time
from sklearn.ensemble import IsolationForest
from sklearn.svm import OneClassSVM
from sklearn.neighbors import LocalOutlierFactor
from sklearn.metrics import confusion_matrix
from sklearn.metrics import roc_auc_score
from matplotlib.mlab import frange
import statistics
from sklearn.metrics import confusion_matrix
from sklearn.metrics import precision_score, recall_score
from sklearn.metrics import f1_score
from sklearn.metrics import roc_curve
```

Chargement du jeu de données KDD99-Cup HTTP

KDD99-Cup HTTP contient 30 attributs dont le dernier est la classe à expliquer. Il y a 620098 observations dont 1052 anormales. Avec le dernier attribut, nous avons les classes "o" pour les anomalies et "n" pour les observations normales.

In [2]:

```
data_brut_KDD99CUP = pd.read_csv('/Users/thesard/Doctorat/These2018/ISEP/Developpements/EspaceIntelliJ/LearningAllInPython/Learning_IHM_Features/datasets/imported_datasets/kdd99-unsupervised-ad_2019-06-24 16:08:29.815617.csv', header=None, index_col=None)
X_KDD99CUP = data_brut_KDD99CUP[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 ,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]]
y_brut_KDD99CUP = data_brut_KDD99CUP[[29]]
data_brut_KDD99CUP.describe()
```

Out[2]:

	0	1	2	3
count	620098.000000	620098.000000	6.200980e+05	620098.000000
mean	0.627670	280.374673	4.099298e+03	0.005012
std	53.972528	1509.901935	2.369388e+04	0.098271
min	0.000000	0.000000	0.000000e+00	0.000000
25%	0.000000	210.000000	4.580000e+02	0.000000
50%	0.000000	240.000000	1.425000e+03	0.000000
75%	0.000000	302.000000	3.731000e+03	0.000000
max	41065.000000	54540.000000	1.173059e+07	21.000000

8 rows × 29 columns

In [3]:

```
# Nombre d'anomalies
n_outliers = data_brut_KDD99CUP[data_brut_KDD99CUP[29] == 'o']
len(n_outliers)/6
```

Out[3]:

```
175.33333333333334
```

In [4]:

```
#Pourcentage d'anomalies
pourcentage_anomalies = len(n_outliers)/len(data_brut_KDD99CUP
)
pourcentage_anomalies
```

Out[4]:

```
0.00169650603614267 43
```

In [5]:

```
# Nombre de données normales
n_normals = data_brut_KDD99CUP[data_brut_KDD99CUP[29] == 'n']
len(n_normals)/6
```

Out[5]:

```
103174.33333333333
```

In [6]:

```
#Pourcentage de données normales
pourcentage_normales = len(n_normals)/len(data_brut_KDD99CUP)
pourcentage_normales
```

Out[6]:

```
0.9983034939638573
```

Spliter le jeu de donner pour réduire la taille

Ceci sera fait en tenant compte du taux d'anomalies dans le jeu de données principale. On aura au final 6 sous jeux de données sur lesquels chaque méthode sera appliquée. La moyenne des résultats sera alors le résultat à retenir pour tout le jeu de données. NB:

```
- CPU Time = Somme
- Précion  = Moyenne ou Somme des différentes matrices
de confusion
- Rappel   = Moyenne ou Somme des différentes matrices
de confusion
- ROC AUC  = Moyenne
- f1-score = Moyenne
- TNBC     = Moyenne ou Somme des différentes matrices
de confusion
- TPBC     = Moyenne ou Somme des différentes matrices
de confusion
```

In [7]:

```
# Spliter à l'aide Stratifield mais ça ne donne pas le résulta
t que je veux. ça donne 6 jeux d'entrainement et
# et de test, tous de même taille dont la somme des jeux de te
st et d'entrainement donne la taille du jeu de
# données original.
#from sklearn.model_selection import StratifiedShuffleSplit
#split = StratifiedShuffleSplit(n_splits=6, test_size=0.169650
60361426743, random_state=42)
#i = 0
#for train_index, test_index in split.split(data_brut_KDD99CUP
, data_brut_KDD99CUP[29]):
#    i=i+1
#    strat_train_set = data_brut_KDD99CUP.loc[train_index]
#    print("train :"+str(i))
#    #print(strat_train_set)
#    print("Taille ="+str(len(strat_train_set)))
#    strat_train_outliers = strat_train_set[strat_train_set[29
] == 'o']
#    strat_train_normals  = strat_train_set[strat_train_set[29
] == 'n']
#    print("n_outliers ="+str(len(strat_train_outliers)/(len(s
trat_train_outliers) + len(strat_train_normals))))
#    print("%_outliers ="+str(len(strat_train_outliers)))
#    print("n_normals  ="+str(len(strat_train_normals)))
#    print("%_normals  ="+str(len(strat_train_normals)/(len(st
rat_train_outliers) + len(strat_train_normals))))
#    strat_test_set = data_brut_KDD99CUP.loc[test_index]
#    print("test :"+str(i))
#    #print(strat_test_set)
#    print("Taille ="+str(len(strat_test_set)))
#    strat_test_outliers = strat_test_set[strat_test_set[29] =
= 'o']
#    strat_test_normals  = strat_test_set[strat_test_set[29] =
= 'n']
#    print("n_outliers ="+str(len(strat_test_outliers)))
#    print("%_outliers ="+str(len(strat_test_outliers)/(len(st
rat_test_outliers) + len(strat_test_normals))))
#    print("n_normals  ="+str(len(strat_test_normals)))
#    print("%_normals  ="+str(len(strat_test_normals)/(len(str
at_test_outliers) + len(strat_test_normals))))
#    print("\n")
```

In [8]:

```
# Subset the dataset by myself
```

```
import math
n_i_min = 0
o_i_min = 0
n_i_max = math.ceil(len(n_normals)/6)
o_i_max = math.ceil(len(n_outliers)/6)
n_sub_dataset_1 = n_normals[n_i_min:n_i_max]
print("Length normals subset 1 = "+str(len(n_sub_dataset_1)))
o_sub_dataset_1 = n_outliers[o_i_min:o_i_max]
print("Length outliers subset 1 = "+str(len(o_sub_dataset_1)))

n_i_min = n_i_max
o_i_min = o_i_max
n_i_max = n_i_max + math.ceil(len(n_normals)/6)
o_i_max = o_i_max + math.ceil(len(n_outliers)/6)
n_sub_dataset_2 = n_normals[n_i_min:n_i_max]
print("Length normals subset 2 = "+str(len(n_sub_dataset_2)))
o_sub_dataset_2 = n_outliers[o_i_min:o_i_max]
print("Length outliers subset 2 = "+str(len(o_sub_dataset_2)))

n_i_min = n_i_max
o_i_min = o_i_max
n_i_max = n_i_max + math.ceil(len(n_normals)/6)
o_i_max = o_i_max + math.ceil(len(n_outliers)/6)
n_sub_dataset_3 = n_normals[n_i_min:n_i_max]
print("Length normals subset 3 = "+str(len(n_sub_dataset_3)))
o_sub_dataset_3 = n_outliers[o_i_min:o_i_max]
print("Length outliers subset 3 = "+str(len(o_sub_dataset_3)))

n_i_min = n_i_max
o_i_min = o_i_max
n_i_max = n_i_max + math.ceil(len(n_normals)/6)
o_i_max = o_i_max + math.ceil(len(n_outliers)/6)
n_sub_dataset_4 = n_normals[n_i_min:n_i_max]
print("Length normals subset 4 = "+str(len(n_sub_dataset_4)))
o_sub_dataset_4 = n_outliers[o_i_min:o_i_max]
print("Length outliers subset 4 = "+str(len(o_sub_dataset_4)))

n_i_min = n_i_max
o_i_min = o_i_max
n_i_max = n_i_max + math.ceil(len(n_normals)/6)
o_i_max = o_i_max + math.ceil(len(n_outliers)/6)
n_sub_dataset_5 = n_normals[n_i_min:n_i_max]
print("Length normals subset 5 = "+str(len(n_sub_dataset_5)))
o_sub_dataset_5 = n_outliers[o_i_min:o_i_max]
print("Length outliers subset 5 = "+str(len(o_sub_dataset_5)))

n_i_min = n_i_max
o_i_min = o_i_max
n_i_max = n_i_max + math.ceil(len(n_normals)/6)
```

```
o_i_max = o_i_max + math.ceil(len(n_outliers)/6)

n_sub_dataset_6 = n_normals[n_i_min:n_i_max]
print("Length normals subset 6 = "+str(len(n_sub_dataset_6)))
o_sub_dataset_6 = n_outliers[o_i_min:o_i_max]
print("Length outliers subset 6 = "+str(len(o_sub_dataset_6)))
```

```
Length normals subset 1 = 103175
Length outliers subset 1 = 176
Length normals subset 2 = 103175
Length outliers subset 2 = 176
Length normals subset 3 = 103175
Length outliers subset 3 = 176
Length normals subset 4 = 103175
Length outliers subset 4 = 176
Length normals subset 5 = 103175
Length outliers subset 5 = 176
Length normals subset 6 = 103171
Length outliers subset 6 = 172
```

Fusion des deux sub_datasets (normales + anomalies)

In [9]:

```
# Fusion of the sub_datasets
sub_dataset_1 = n_sub_dataset_1.append(o_sub_dataset_1, ignore
_index=True, sort=False)
sub_dataset_2 = n_sub_dataset_2.append(o_sub_dataset_2, ignore
_index=True, sort=False)
sub_dataset_3 = n_sub_dataset_3.append(o_sub_dataset_3, ignore
_index=True, sort=False)
sub_dataset_4 = n_sub_dataset_4.append(o_sub_dataset_4, ignore
_index=True, sort=False)
sub_dataset_5 = n_sub_dataset_5.append(o_sub_dataset_5, ignore
_index=True, sort=False)
sub_dataset_6 = n_sub_dataset_6.append(o_sub_dataset_6, ignore
_index=True, sort=False)
#sub_dataset_6.describe()
# Détacher la classe à expliquer des autres classes
X_sub_dataset_1 = sub_dataset_1[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12, 13 ,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
, 26, 27, 28]]
y_sub_dataset_1 = sub_dataset_1[[29]]
X_sub_dataset_2 = sub_dataset_2[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12, 13 ,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
, 26, 27, 28]]
y_sub_dataset_2 = sub_dataset_2[[29]]
X_sub_dataset_3 = sub_dataset_3[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12, 13 ,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
, 26, 27, 28]]
y_sub_dataset_3 = sub_dataset_3[[29]]
X_sub_dataset_4 = sub_dataset_4[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12, 13 ,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
, 26, 27, 28]]
y_sub_dataset_4 = sub_dataset_4[[29]]
X_sub_dataset_5 = sub_dataset_5[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12, 13 ,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
, 26, 27, 28]]
y_sub_dataset_5 = sub_dataset_5[[29]]
X_sub_dataset_6 = sub_dataset_6[[0, 1, 2, 3, 4, 5, 6, 7, 8, 9,
10, 11, 12, 13 ,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
, 26, 27, 28]]
y_sub_dataset_6 = sub_dataset_6[[29]]
```

Transformation de la classe à expliquer

L'objectif ici est de remplacer les "o" par -1 et les "n" par 1 afin de faire les matrices de confusion avec la fonction dédiée de scikit-learn.

In [10]:

```
y_transform_KDD99CUP       = y_brut_KDD99CUP
y_transform_KDD99CUP       = np.where(y_transform_KDD99CUP =='o
',-1,1)
#y_transform_KDD99CUP

y_transform_sub_dataset_6 = np.where(y_sub_dataset_6 =='o',-1,
1)
y_transform_sub_dataset_5 = np.where(y_sub_dataset_5 =='o',-1,
1)
y_transform_sub_dataset_4 = np.where(y_sub_dataset_4 =='o',-1,
1)
y_transform_sub_dataset_3 = np.where(y_sub_dataset_3 =='o',-1,
1)
y_transform_sub_dataset_2 = np.where(y_sub_dataset_2 =='o',-1,
1)
y_transform_sub_dataset_1 = np.where(y_sub_dataset_1 =='o',-1,
1)
```

KDD99CUP

Exécution de Isolation Forest sur le jeux de données KDD99CUP.

In [11]:

```
max_samples = 'auto'

averages_executions_time_IF_KDD99CUP = []
averages_roc_auc_IF_KDD99CUP = []

for j in range(10, 300, 10):
    n_estimators = j
    executions_time_IF_KDD99CUP = []
    roc_auc_IF_KDD99CUP = []
    #for i in range(5):
############################
    roc_auc_IF_Sub_Datasets = []
    start_IF_KDD99CUP = time()
```

```
    # Apply the function on the 6 sub_dataset

    func_IF_KDD99CUP = IsolationForest(n_estimators=n_estimato
rs, max_samples=max_samples)
    func_IF_KDD99CUP.fit(X_sub_dataset_1)
    y_pred_IF_sub_dataset_1 = func_IF_KDD99CUP.predict(X_sub_d
ataset_1)

    #func_IF_KDD99CUP = IsolationForest(n_estimators=n_estimat
ors, max_samples=max_samples)
    #func_IF_KDD99CUP.fit(X_sub_dataset_2)
    #y_pred_IF_sub_dataset_2 = func_IF_KDD99CUP.predict(X_sub_
dataset_2)

    #func_IF_KDD99CUP = IsolationForest(n_estimators=n_estimat
ors, max_samples=max_samples)
    #func_IF_KDD99CUP.fit(X_sub_dataset_3)
    #y_pred_IF_sub_dataset_3 = func_IF_KDD99CUP.predict(X_sub_
dataset_3)

    #func_IF_KDD99CUP = IsolationForest(n_estimators=n_estimat
ors, max_samples=max_samples)
    #func_IF_KDD99CUP.fit(X_sub_dataset_4)
    #y_pred_IF_sub_dataset_4 = func_IF_KDD99CUP.predict(X_sub_
dataset_4)

    #func_IF_KDD99CUP = IsolationForest(n_estimators=n_estimat
ors, max_samples=max_samples)
    #func_IF_KDD99CUP.fit(X_sub_dataset_5)
    #y_pred_IF_sub_dataset_5 = func_IF_KDD99CUP.predict(X_sub_
dataset_5)

    #func_IF_KDD99CUP = IsolationForest(n_estimators=n_estimat
ors, max_samples=max_samples)
    #func_IF_KDD99CUP.fit(X_sub_dataset_6)
    #y_pred_IF_sub_dataset_6 = func_IF_KDD99CUP.predict(X_sub_
dataset_6)

    # Calcul du temps d'exécution
    exec_time_IF_KDD99CUP = time() - start_IF_KDD99CUP
    executions_time_IF_KDD99CUP.append(exec_time_IF_KDD99CUP)

    # Calcul de l'aire sous la courbe ROC
    # Apply the function on ROC AUC on the 6 sub_dataset
    auc_IF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_1,
y_pred_IF_sub_dataset_1)
    roc_auc_IF_Sub_Datasets.append(auc_IF_KDD99CUP)
    #auc_IF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_2
, y_pred_IF_sub_dataset_2)
```

```
    #roc_auc_IF_Sub_Datasets.append(auc_IF_KDD99CUP)

    #auc_IF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_3
, y_pred_IF_sub_dataset_3)
    #roc_auc_IF_Sub_Datasets.append(auc_IF_KDD99CUP)
    #auc_IF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_4
, y_pred_IF_sub_dataset_4)
    #roc_auc_IF_Sub_Datasets.append(auc_IF_KDD99CUP)
    #auc_IF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_5
, y_pred_IF_sub_dataset_5)
    #roc_auc_IF_Sub_Datasets.append(auc_IF_KDD99CUP)
    #auc_IF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_6
, y_pred_IF_sub_dataset_6)
    #roc_auc_IF_Sub_Datasets.append(auc_IF_KDD99CUP)

    # Mean of the ROC AUC of the 6 sub_datasets
    roc_auc_IF_KDD99CUP.append(statistics.mean(roc_auc_IF_Sub_
Datasets))
    #print("ROC AUC for Isolation Forest applied on KDD99CUP")
    #print(roc_auc_IF_KDD99CUP)
 ########################################
    print("Execution time of IsolationForest applied on KDD99C
UP (in second)")
    print(executions_time_IF_KDD99CUP)
    print("ROC AUC for Isolation Forest applied on KDD99CUP")
    print(roc_auc_IF_KDD99CUP)
    averages_executions_time_IF_KDD99CUP.append(statistics.mea
n(executions_time_IF_KDD99CUP))
    averages_roc_auc_IF_KDD99CUP.append(statistics.mean(roc_au
c_IF_KDD99CUP))
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[1.005812168121338]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[1.8286850452423096]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9450766019781042]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[2.8137331008911133]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[3.6294209957122803]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[4.574329137802124]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9422308467519 88]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[5.406002044677734]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[6.241639137268066]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[8.099387884140015]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9450766019781042]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[10.536028146743774]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[9.836879968643188]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[12.066686153411865]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.94792235720422O5]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[12.566919088363647]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[11.864881038665771]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9393850915258719]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[13.087165832519531]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[14.374619960784912]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[15.394497871398926]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[16.392831087112427]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[15.989430665969849]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.942230846751988]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[16.98485493659973]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[17.7335278987884 52]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[18.768343925476074]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[20.13857364654541]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[20.72728991508484]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[26.03485608100891]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[21.59841823577881]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.950772958565544]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[22.76730990409851]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)


Execution time of IsolationForest applied on KDD99
CUP (in second)
[23.674289226531982]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)

Execution time of IsolationForest applied on KDD99
CUP (in second)
[26.112324953079224]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9508068815119941]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
```

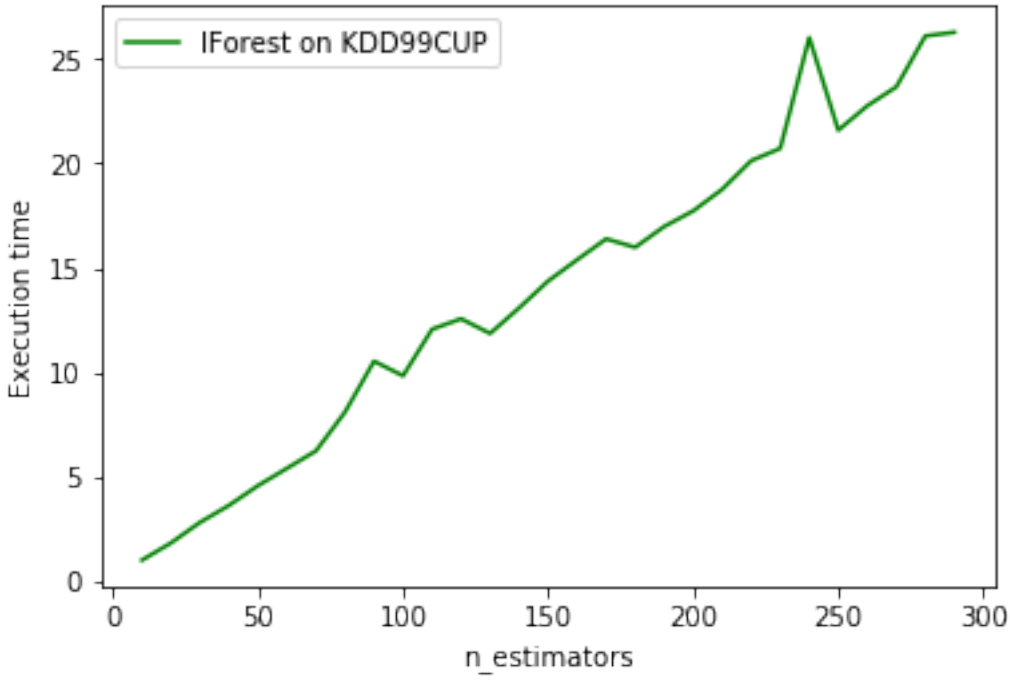
```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[26.287743091583252]
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369]
```

Résultat de Isolation Forest sur KDD99CUP

In [12]:

```
print("Execution time of IsolationForest applied on KDD99CUP (
in second)")
print(averages_executions_time_IF_KDD99CUP)
plt.plot(range(10, 300, 10), averages_executions_time_IF_KDD99
CUP,  "g-", label="IForest on KDD99CUP")
#plt.axis([0, 1, 0, 1])
plt.xlabel('n_estimators')
plt.ylabel('Execution time')
plt.legend(loc="best")
plt.show()
```

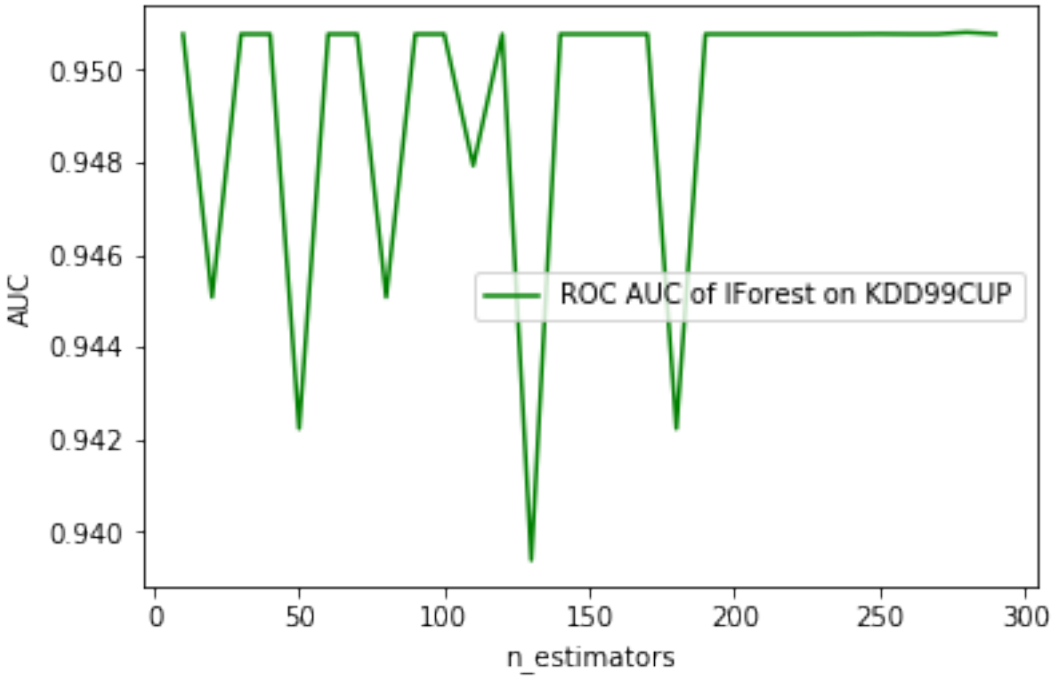
```
Execution time of IsolationForest applied on KDD99
CUP (in second)
[1.005812168121338, 1.828685045242309 6, 2.81373310
08911133, 3.6294209957122803, 4.574329137802124, 5
.406002044677734, 6.241639137268066, 8.09938788414
0015, 10.536028146743774, 9.836879968643188, 12.06
6686153411865, 12.566919088363647, 11.864881038665
771, 13.087165832519531, 14.374619960784912, 15.39
4497871398926, 16.392831087112427, 15.989430665969
849, 16.98485493659973, 17.733527898788452, 18.768
343925476074, 20.13857364654541, 20.72728991508484
, 26.03485608100891, 21.59841823577881, 22.7673099
0409851, 23.674289226531982, 26.112324953079224, 2
6.287743091583252]
```



In [13]:

```
print("ROC AUC for Isolation Forest applied on KDD99CUP")
print(averages_roc_auc_IF_KDD99CUP)
print("La valeur max est "+str(max(averages_roc_auc_IF_KDD99CU
P))+" à l'index "+str(averages_roc_auc_IF_KDD99CUP.index(max(a
verages_roc_auc_IF_KDD99CUP))))
plt.plot(range(10, 300, 10), averages_roc_auc_IF_KDD99CUP,  "g
-", label="ROC AUC of IForest on KDD99CUP")
#plt.axis([0, 1, 0, 1])
plt.xlabel('n_estimators')
plt.ylabel('AUC')
plt.legend(loc="best")
plt.show()
```

```
ROC AUC for Isolation Forest applied on KDD99CUP
[0.9507681124303369, 0.9450766019781042, 0.9507681
124303369, 0.9507681124303369, 0.942230846751988,
0.9507681124303369, 0.9507681124303369, 0.94507660
19781042, 0.9507681124303369, 0.9507681124303369,
0.9479223572042205, 0.9507681124303369, 0.93938509
15258719, 0.9507681124303369, 0.9507681124303369,
0.9507681124303369, 0.9507681124303369, 0.94223084
6751988, 0.9507681124303369, 0.9507681124303369, 0
.9507681124303369, 0.9507681124303369, 0.950768112
4303369, 0.9507681124303369, 0.950772958565544, 0.
9507681124303369, 0.9507681124303369, 0.9508068815
119941, 0.9507681124303369]
La valeur max est 0.9508068815119941 à l'index 27
```



Interprétation du résultat de IsolationForest sur KDD99CUP

Avec n_estimators entre 10 et 300, on remarque que le temps d'exécution augmente au fur et à mesure que le nombre d'estimateurs augmente. Concernant la performance, la meilleure estimation a été obtenu pour un taux de AUC = 0.0.9508068815119941 avec n_estimators = 280 ===> (10+27*10) A noter que quelque soit la valeur de n_estimators, la AUC est toujours supérieure à 0,94

Peut-être qu'au delà de 300 estimateurs, la performance augmentera

Exécution de LOF sur KDD99CUP

In [14]:

```
averages_executions_time_LOF_KDD99CUP = []
averages_roc_auc_LOF_KDD99CUP = []

for j in range(5, 200, 5):
    n_neighbors = j
    executions_time_LOF_KDD99CUP = []
    roc_auc_LOF_KDD99CUP = []

    #for i in range(5):
############################
    roc_auc_LOF_Sub_Datasets = []
    start_LOF_KDD99CUP = time()

    # Apply the function on the 6 sub_dataset
    func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neigh
bors,novelty=True)
    func_LOF_KDD99CUP.fit(X_sub_dataset_1)
    y_pred_LOF_sub_dataset_1 = func_LOF_KDD99CUP.predict(X_sub
_dataset_1)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_2)
    #y_pred_LOF_sub_dataset_2 = func_LOF_KDD99CUP.predict(X_su
b_dataset_2)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_3)
    #y_pred_LOF_sub_dataset_3 = func_LOF_KDD99CUP.predict(X_su
b_dataset_3)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_4)
    #y_pred_LOF_sub_dataset_4 = func_LOF_KDD99CUP.predict(X_su
b_dataset_4)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_5)
    #y_pred_LOF_sub_dataset_5 = func_LOF_KDD99CUP.predict(X_su
b_dataset_5)
```

```
    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neighbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_6)
    #y_pred_LOF_sub_dataset_6 = func_LOF_KDD99CUP.predict(X_sub_dataset_6)

    # Calcul du temps d'exécution
    exec_time_LOF_KDD99CUP = time() - start_LOF_KDD99CUP
    executions_time_LOF_KDD99CUP.append(exec_time_LOF_KDD99CUP)

    # Calcul de l'aire sous la courbe ROC
    # Apply the function on ROC AUC on the 6 sub_dataset
    auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_1, y_pred_LOF_sub_dataset_1)
    roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_2, y_pred_LOF_sub_dataset_2)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_3, y_pred_LOF_sub_dataset_3)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_4, y_pred_LOF_sub_dataset_4)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_5, y_pred_LOF_sub_dataset_5)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_6, y_pred_LOF_sub_dataset_6)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)

    # Mean of the ROC AUC of the 6 sub_datasets
    roc_auc_LOF_KDD99CUP.append(statistics.mean(roc_auc_LOF_Sub_Datasets))
 ##########################################

    print("Execution time of LOF applied on KDD99CUP (in second)")
    print(executions_time_LOF_KDD99CUP)
    print("ROC AUC for LOF applied on KDD99CUP")
    print(roc_auc_LOF_KDD99CUP)
    averages_executions_time_LOF_KDD99CUP.append(statistics.mean(executions_time_LOF_KDD99CUP))
    averages_roc_auc_LOF_KDD99CUP.append(statistics.mean(roc_auc_LOF_KDD99CUP))
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[18.902546882629395]
ROC AUC for LOF applied on KDD99CUP
[0.5477257307751613]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[22.910732984542847]
ROC AUC for LOF applied on KDD99CUP
[0.5563825252770007]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[25.524651050567627]
ROC AUC for LOF applied on KDD99CUP
[0.5417089785668656]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[28.405873775482178]
ROC AUC for LOF applied on KDD99CUP
[0.52326783708174 55]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[24.597292184829712]
ROC AUC for LOF applied on KDD99CUP
[0.5024025541335331]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[26.512227058410645]
ROC AUC for LOF applied on KDD99CUP
[0.49890257065444854]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[30.136018991470337]
ROC AUC for LOF applied on KDD99CUP
[0.49869903297574 73]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[27.56598997116089]
ROC AUC for LOF applied on KDD99CUP
[0.4979818049650858]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[29.637439966201782]
ROC AUC for LOF applied on KDD99CUP
[0.4979575742890499 6]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[33.470175981521606]
ROC AUC for LOF applied on KDD99CUP
[0.4975844218780977]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[33.769304037094116]
ROC AUC for LOF applied on KDD99CUP
[0.5256785139987223]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[34.111266136169434]
ROC AUC for LOF applied on KDD99CUP
[0.5427433530850057]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[30.4944269657135]
ROC AUC for LOF applied on KDD99CUP
[0.5940784082648634]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[34.438093185424805]
ROC AUC for LOF applied on KDD99CUP
[0.6052530453554199]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[38.32731628417969]
ROC AUC for LOF applied on KDD99CUP
[0.5965073958631628]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[40.01055288314819]
ROC AUC for LOF applied on KDD99CUP
[0.6021455988281164]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[34.12298798561096]
ROC AUC for LOF applied on KDD99CUP
[0.6047732779699098]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[36.494633197784424]
ROC AUC for LOF applied on KDD99CUP
[0.5677929984360199]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[39.581904888153076]
ROC AUC for LOF applied on KDD99CUP
[0.5506554673216292]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[41.35542297363281]
ROC AUC for LOF applied on KDD99CUP
[0.5448524957596317]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[40.207817792892456]
ROC AUC for LOF applied on KDD99CUP
[0.5391803698482278]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[42.744410276412964]
ROC AUC for LOF applied on KDD99CUP
[0.5304686433024208]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[43.08960795402527]
ROC AUC for LOF applied on KDD99CUP
[0.5304347203559707]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[43.80181670188904]
ROC AUC for LOF applied on KDD99CUP
[0.5191437760204419]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[39.523130893707275]
ROC AUC for LOF applied on KDD99CUP
[0.5132971892415799]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[43.94402503967285]
ROC AUC for LOF applied on KDD99CUP
[0.513258420159922 4]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[43.29129695892334]
ROC AUC for LOF applied on KDD99CUP
[0.5075281406260326]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[47.453534841537476]
ROC AUC for LOF applied on KDD99CUP
[0.4874673161222107]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[46.82526993751526]
ROC AUC for LOF applied on KDD99CUP
[0.4874721622574179]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[47.25864815711975]
ROC AUC for LOF applied on KDD99CUP
[0.48467486838337337]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[46.17681694030762]
ROC AUC for LOF applied on KDD99CUP
[0.4846021763552658]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[46.768396615982056]
ROC AUC for LOF applied on KDD99CUP
[0.484524638191951]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[53.82866930961609]
ROC AUC for LOF applied on KDD99CUP
[0.484544022732779 74]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[57.695087909698486]
ROC AUC for LOF applied on KDD99CUP
[0.4930909806815428 3]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[54.3389208316803]
ROC AUC for LOF applied on KDD99CUP
[0.4931249036279930 5]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[52.49622702598572]
ROC AUC for LOF applied on KDD99CUP
[0.4959270436372447 7]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[49.82447576522827]
ROC AUC for LOF applied on KDD99CUP
[0.4960191202061810 5]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[50.597997188568115]
ROC AUC for LOF applied on KDD99CUP
[0.516026637222724]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

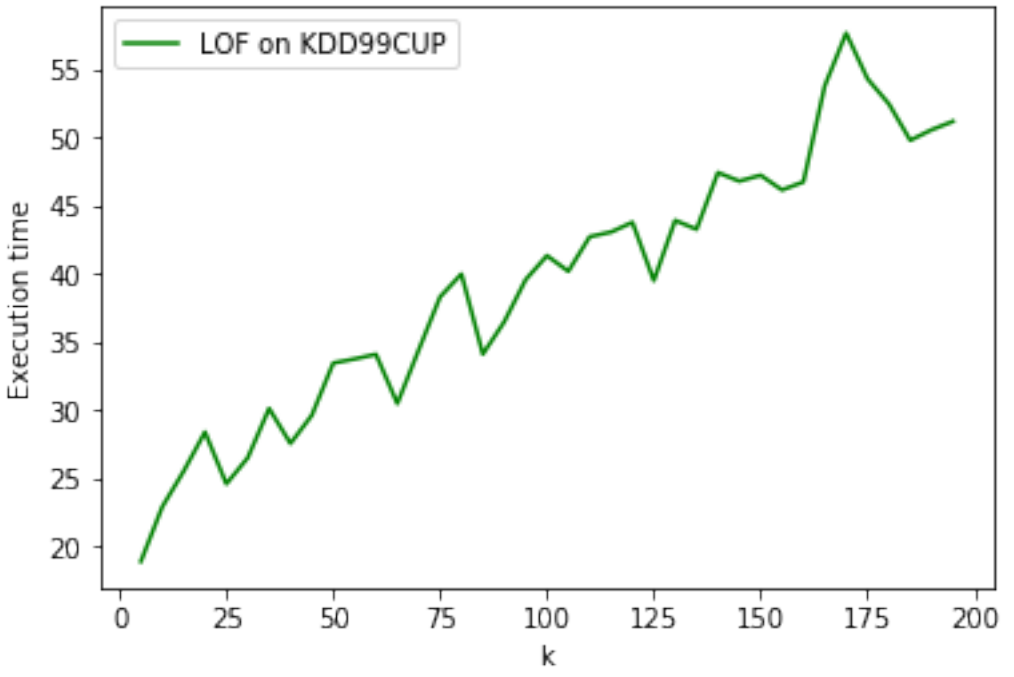
```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[51.21554613113403]
ROC AUC for LOF applied on KDD99CUP
[0.783396782827059]
```

Résultat de LOF sur KDD99CUP

In [15]:

```
print("Execution time of LOF applied on KDD99CUP (in second)")
print(averages_executions_time_LOF_KDD99CUP)
plt.plot(range(5, 200, 5), averages_executions_time_LOF_KDD99C
UP,  "g-", label="LOF on KDD99CUP")
#plt.axis([0, 200, 0, 200])
plt.xlabel('k')
plt.ylabel('Execution time')
plt.legend(loc="best")
plt.show()
```

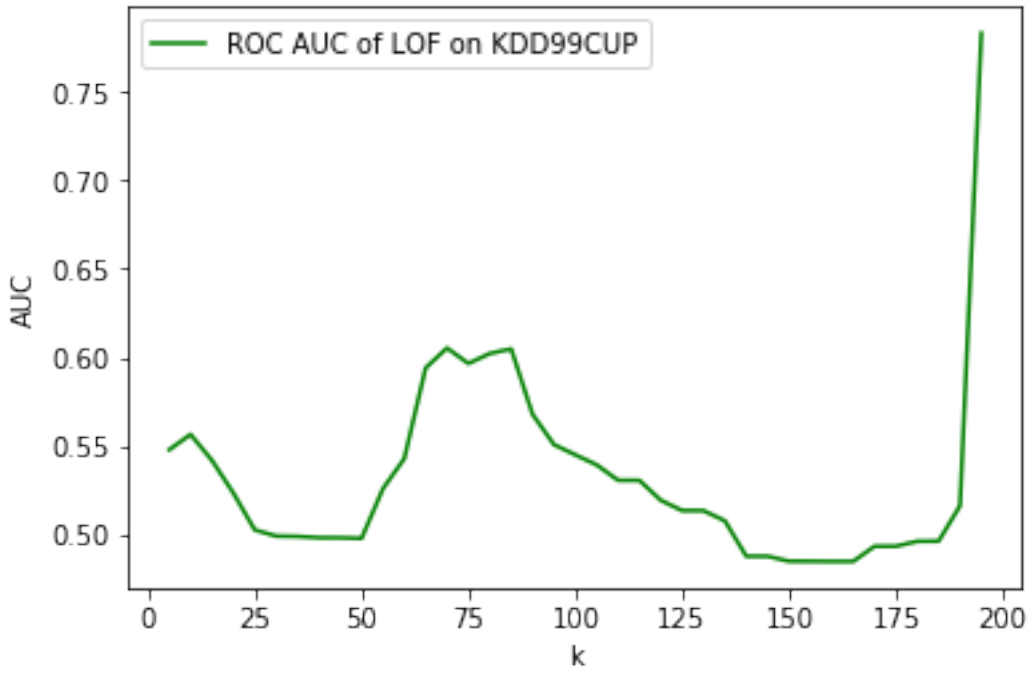
Execution time of LOF applied on KDD99CUP (in seco
nd)
[18.902546882629395, 22.910732984542847, 25.524651
050567627, 28.405873775482178, 24.597292184829712,
26.512227058410645, 30.136018991470337, 27.5659899
7116089, 29.637439966201782, 33.470175981521606, 3
3.769304037094116, 34.111266136169434, 30.49442696
57135, 34.438093185424805, 38.32731628417969, 40.0
1055288314819, 34.12298798561096, 36.4946331977844
24, 39.581904888153076, 41.35542297363281, 40.2078
17792892456, 42.744410276412964, 43.08960795402527
, 43.80181670188904, 39.523130893707275, 43.944025
03967285, 43.29129695892334, 47.453534841537476, 4
6.82526993751526, 47.25864815711975, 46.1768169403
0762, 46.768396615982056, 53.82866930961609, 57.69
5087909698486, 54.3389208316803, 52.49622702598572
, 49.82447576522827, 50.597997188568115, 51.215546
13113403]



In [16]:

```
print("ROC AUC for LOF applied on KDD99CUP")
print(averages_roc_auc_LOF_KDD99CUP)
print("La valeur max est "+str(max(averages_roc_auc_LOF_KDD99C
UP))+" à l'index "+str(averages_roc_auc_LOF_KDD99CUP.index(max
(averages_roc_auc_LOF_KDD99CUP))))
plt.plot(range(5, 200, 5), averages_roc_auc_LOF_KDD99CUP,  "g-
", label="ROC AUC of LOF on KDD99CUP")
#plt.axis([0, 1, 0, 1])
plt.xlabel('k')
plt.ylabel('AUC')
plt.legend(loc="best")
plt.show()
```

```
ROC AUC for LOF applied on KDD99CUP
[0.5477257307751613, 0.5563825252770007, 0.5417089
785668656, 0.5232678370817455, 0.5024025541335331,
0.4989025706544854, 0.4986990329757473, 0.4979818
049650858, 0.4979575742890496, 0.4975844218780977
, 0.5256785139987223, 0.5427433530850057, 0.594078
4082648634, 0.6052530453554199, 0.5965073958631628
, 0.6021455988281164, 0.6047732779699098, 0.567792
9984360199, 0.5506554673216292, 0.5448524957596317
, 0.5391803698482278, 0.5304686433024208, 0.530434
7203559707, 0.5191437760204419, 0.5132971892415799
, 0.5132584201599224, 0.5075281406260326, 0.487467
3161222107, 0.4874721622574179, 0.4846748683833733
7, 0.4846021763552658, 0.484524638191951, 0.484544
02273277974, 0.49309098068154283, 0.49312490362799
305, 0.49592704363724477, 0.49601912020618105, 0.5
16026637222724, 0.783396782827059]
La valeur max est 0.783396782827059 à l'index 38
```



In [20]:

```
averages_executions_time_LOF_KDD99CUP = []
averages_roc_auc_LOF_KDD99CUP = []

for j in range(5, 200, 5):
    n_neighbors = j
    executions_time_LOF_KDD99CUP = []
    roc_auc_LOF_KDD99CUP = []

    #for i in range(5):
############################
    roc_auc_LOF_Sub_Datasets = []
```

```
    start_LOF_KDD99CUP = time()

    # Apply the function on the 6 sub_dataset
    func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neigh
bors)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_1)
    y_pred_LOF_sub_dataset_1 = func_LOF_KDD99CUP.fit_predict(X
_sub_dataset_1)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_2)
    #y_pred_LOF_sub_dataset_2 = func_LOF_KDD99CUP.predict(X_su
b_dataset_2)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_3)
    #y_pred_LOF_sub_dataset_3 = func_LOF_KDD99CUP.predict(X_su
b_dataset_3)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_4)
    #y_pred_LOF_sub_dataset_4 = func_LOF_KDD99CUP.predict(X_su
b_dataset_4)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_5)
    #y_pred_LOF_sub_dataset_5 = func_LOF_KDD99CUP.predict(X_su
b_dataset_5)

    #func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neig
hbors,novelty=True)
    #func_LOF_KDD99CUP.fit(X_sub_dataset_6)
    #y_pred_LOF_sub_dataset_6 = func_LOF_KDD99CUP.predict(X_su
b_dataset_6)

    # Calcul du temps d'exécution
    exec_time_LOF_KDD99CUP = time() - start_LOF_KDD99CUP
    executions_time_LOF_KDD99CUP.append(exec_time_LOF_KDD99CUP
)

    # Calcul de l'aire sous la courbe ROC
    # Apply the function on ROC AUC on the 6 sub_dataset
    auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_1
, y_pred_LOF_sub_dataset_1)
```

```
    roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)

    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_
2, y_pred_LOF_sub_dataset_2)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_
3, y_pred_LOF_sub_dataset_3)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_
4, y_pred_LOF_sub_dataset_4)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_
5, y_pred_LOF_sub_dataset_5)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)
    #auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_
6, y_pred_LOF_sub_dataset_6)
    #roc_auc_LOF_Sub_Datasets.append(auc_LOF_KDD99CUP)

    # Mean of the ROC AUC of the 6 sub_datasets
    roc_auc_LOF_KDD99CUP.append(statistics.mean(roc_auc_LOF_Su
b_Datasets))
 ##########################################

    print("Execution time of LOF applied on KDD99CUP (in secon
d)")
    print(executions_time_LOF_KDD99CUP)
    print("ROC AUC for LOF applied on KDD99CUP")
    print(roc_auc_LOF_KDD99CUP)
    averages_executions_time_LOF_KDD99CUP.append(statistics.me
an(executions_time_LOF_KDD99CUP))
    averages_roc_auc_LOF_KDD99CUP.append(statistics.mean(roc_a
uc_LOF_KDD99CUP))
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[16.601251125335693]
ROC AUC for LOF applied on KDD99CUP
[0.5779741778091063]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[17.093443155288696]
ROC AUC for LOF applied on KDD99CUP
[0.5608996464524088]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[18.05699586868286]
ROC AUC for LOF applied on KDD99CUP
[0.5352878494173623]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[18.527397871017456]
ROC AUC for LOF applied on KDD99CUP
[0.521059073286781]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[19.360121965408325]
ROC AUC for LOF applied on KDD99CUP
[0.49829303147785 09]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[23.05241298675537]
ROC AUC for LOF applied on KDD99CUP
[0.49544727625173474]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[24.409388065338135]
ROC AUC for LOF applied on KDD99CUP
[0.49544727625173474]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[23.55992817878723]
ROC AUC for LOF applied on KDD99CUP
[0.49544727625173474]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[22.666385889053345]
ROC AUC for LOF applied on KDD99CUP
[0.49544727625173474]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[23.743576765060425]
ROC AUC for LOF applied on KDD99CUP
[0.49544727625173474]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[24.932187795639038]
ROC AUC for LOF applied on KDD99CUP
[0.52390482851289 73]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[24.036752939224243]
ROC AUC for LOF applied on KDD99CUP
[0.5409793598695949]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[24.62563681602478]
ROC AUC for LOF applied on KDD99CUP
[0.6035859748441527]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[25.987552881240845]
ROC AUC for LOF applied on KDD99CUP
[0.6121232405225014]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[25.4359290599823]
ROC AUC for LOF applied on KDD99CUP
[0.6035859748441527]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[25.830070972442627]
ROC AUC for LOF applied on KDD99CUP
[0.6064317300702 69]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[27.105626106262207]
ROC AUC for LOF applied on KDD99CUP
[0.6064317300702 69]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[28.707056045532227]
ROC AUC for LOF applied on KDD99CUP
[0.6007402196180365]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[27.07657289505005]
ROC AUC for LOF applied on KDD99CUP
[0.5580538912262925]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[27.514561891555786]
ROC AUC for LOF applied on KDD99CUP
[0.5438251150957112]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[27.310483932495117]
ROC AUC for LOF applied on KDD99CUP
[0.5381336046434786]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[27.94042682647705]
ROC AUC for LOF applied on KDD99CUP
[0.5295963389651299]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[28.835601806640625]
ROC AUC for LOF applied on KDD99CUP
[0.529601185100337]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[28.472607851028442]
ROC AUC for LOF applied on KDD99CUP
[0.5182133180606647]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[28.879562854766846]
ROC AUC for LOF applied on KDD99CUP
[0.5125218076084324]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[29.87860131263733]
ROC AUC for LOF applied on KDD99CUP
[0.51252180760843 24]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[29.516635179519653]
ROC AUC for LOF applied on KDD99CUP
[0.5068302971561998]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[32.980283975601196]
ROC AUC for LOF applied on KDD99CUP
[0.48691001057338 59]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[30.841774702072144]
ROC AUC for LOF applied on KDD99CUP
[0.48691001057338 59]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[30.897964000701904]
ROC AUC for LOF applied on KDD99CUP
[0.48406425534726966]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[31.554543018341064]
ROC AUC for LOF applied on KDD99CUP
[0.48406425534726966]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[32.44221019744873]
ROC AUC for LOF applied on KDD99CUP
[0.48406425534726966]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
```

```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[32.38812708854675]
ROC AUC for LOF applied on KDD99CUP
[0.48406425534726966]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[33.433014154434204]
ROC AUC for LOF applied on KDD99CUP
[0.4926015210256185]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[32.84754705429077]
ROC AUC for LOF applied on KDD99CUP
[0.4926015210256185]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[32.687320709228516]
ROC AUC for LOF applied on KDD99CUP
[0.49544727625173474]
```

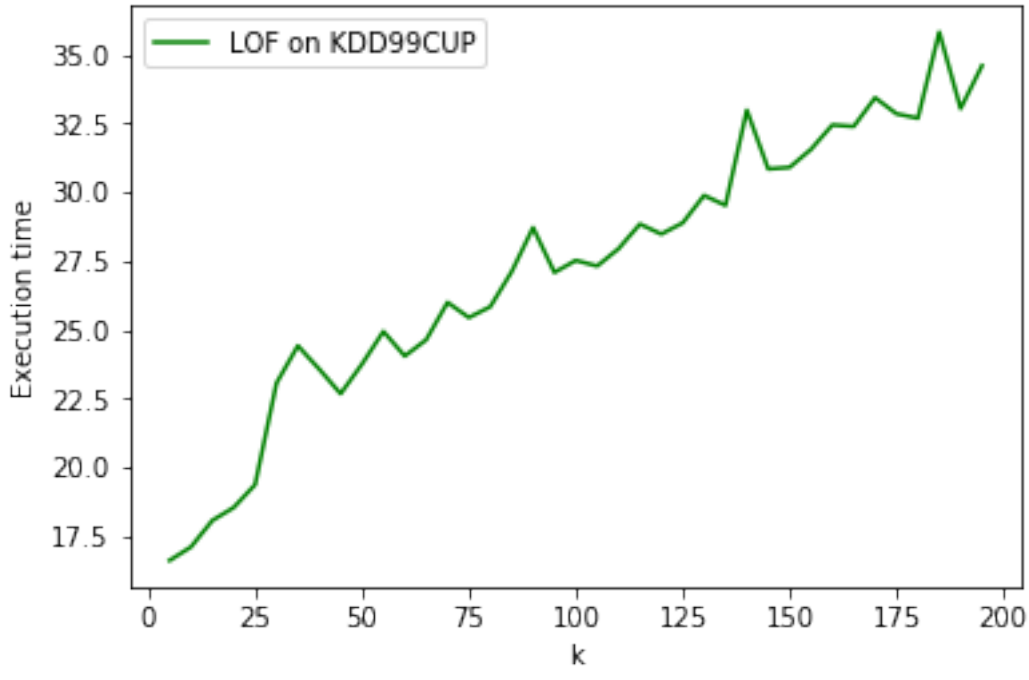
```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)


Execution time of LOF applied on KDD99CUP (in seco
nd)
[35.819053173065186]
ROC AUC for LOF applied on KDD99CUP
[0.49544727625173474]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[33.037994146347046]
ROC AUC for LOF applied on KDD99CUP
[0.5779741778091063]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)

Execution time of LOF applied on KDD99CUP (in seco
nd)
[34.59910607337952]
ROC AUC for LOF applied on KDD99CUP
[0.7885600645417098]
```

Résultat de LOF sur KDD99CUP

In [21]:

```
print("Execution time of LOF applied on KDD99CUP (in second)")
print(averages_executions_time_LOF_KDD99CUP)
plt.plot(range(5, 200, 5), averages_executions_time_LOF_KDD99C
UP,  "g-", label="LOF on KDD99CUP")
#plt.axis([0, 200, 0, 200])
plt.xlabel('k')
plt.ylabel('Execution time')
plt.legend(loc="best")
plt.show()
```

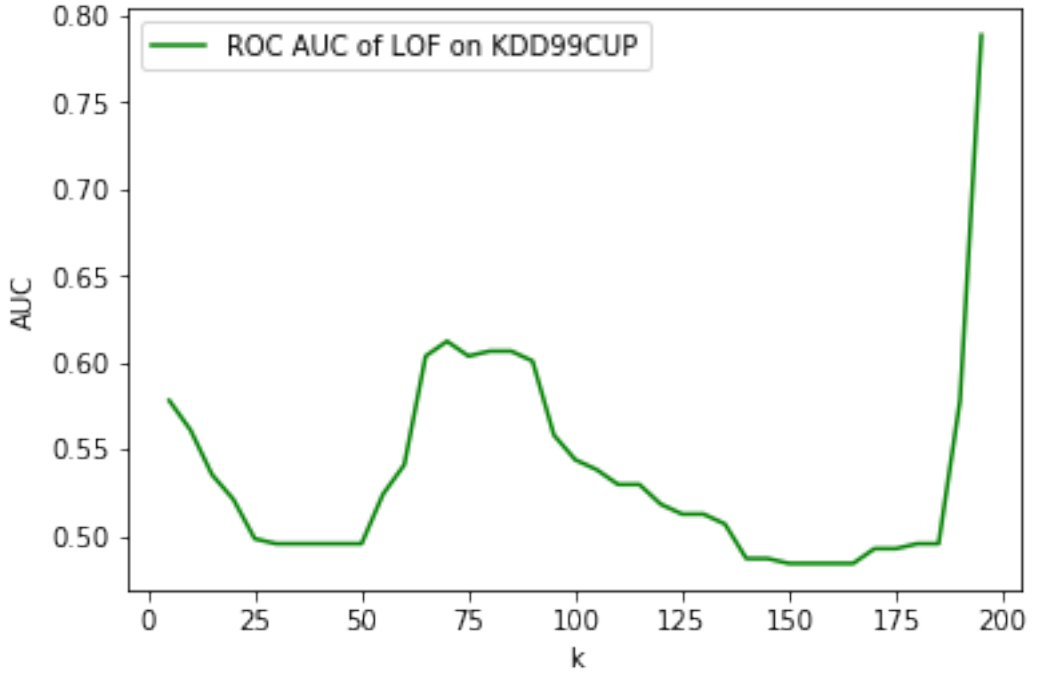
```
Execution time of LOF applied on KDD99CUP (in seco
nd)
[16.601251125335693, 17.093443155288696, 18.056995
86868286, 18.527397871017456, 19.360121965408325,
23.05241298675537, 24.409388065338135, 23.55992817
878723, 22.666385889053345, 23.743576765060425, 24
.932187795639038, 24.036752939224243, 24.625636816
02478, 25.987552881240845, 25.4359290599823, 25.83
0070972442627, 27.105626106262207, 28.707056045532
227, 27.07657289505005, 27.514561891555786, 27.310
483932495117, 27.94042682647705, 28.83560180664062
5, 28.472607851028442, 28.879562854766846, 29.8786
0131263733, 29.516635179519653, 32.980283975601196
, 30.841774702072144, 30.897964000701904, 31.55454
3018341064, 32.44221019744873, 32.38812708854675,
33.433014154434204, 32.84754705429077, 32.68732070
9228516, 35.819053173065186, 33.037994146347046, 3
4.59910607337952]
```



In [22]:

```
print("ROC AUC for LOF applied on KDD99CUP")
print(averages_roc_auc_LOF_KDD99CUP)
print("La valeur max est "+str(max(averages_roc_auc_LOF_KDD99C
UP))+" à l'index "+str(averages_roc_auc_LOF_KDD99CUP.index(max
(averages_roc_auc_LOF_KDD99CUP))))
plt.plot(range(5, 200, 5), averages_roc_auc_LOF_KDD99CUP,  "g-
", label="ROC AUC of LOF on KDD99CUP")
#plt.axis([0, 1, 0, 1])
plt.xlabel('k')
plt.ylabel('AUC')
plt.legend(loc="best")
plt.show()
```

```
ROC AUC for LOF applied on KDD99CUP
[0.5779741778091063, 0.5608996464524088, 0.5352878
494173623, 0.521059073286781, 0.4982930314778509,
0.49544727625173474, 0.49544727625173474, 0.495447
27625173474, 0.49544727625173474, 0.49544727625173
474, 0.5239048285128973, 0.5409793598695949, 0.603
5859748441527, 0.6121232405225014, 0.6035859748441
527, 0.606431730070269, 0.606431730070269, 0.60074
02196180365, 0.5580538912262925, 0.543825115095711
2, 0.5381336046434786, 0.5295963389651299, 0.52960
1185100337, 0.5182133180606647, 0.5125218076084324
, 0.5125218076084324, 0.5068302971561998, 0.486910
0105733859, 0.4869100105733859, 0.4840642553472696
6, 0.48406425534726966, 0.48406425534726966, 0.484
06425534726966, 0.4926015210256185, 0.492601521025
6185, 0.49544727625173474, 0.49544727625173474, 0.
5779741778091063, 0.7885600645417098]
La valeur max est 0.7885600645417098 à l'index 38
```



Interprétation du résultat de LOF sur KDD99CUP

Plus le nombre de n_neighbors augmente, plus de temps d'éxécution est important. Ceci s'explique par le fait que plus le nombre de voisins à considérer est grand plus il faudra calculer les distances et les densité afin d'estimer le degré d'abberrance d'une observation. Concernant la performance de la méthode, elle atteint son plus fort (AUC = 0.7885600645417098) à taux à n_neighbors= 195 ((38+1)*5) pour un temps d'exécution de 405.32807874679565 NB: Continuer l'expérience en prenant n_neighbors à partir de 200 jusqu'à 400 pour voir si ça continuera à mieux classer.

Exécution de OC-SVM sur KDD99CUP

In [17]:

```
averages_executions_time_OCSVM_KDD99CUP = []
averages_roc_auc_OCSVM_KDD99CUP = []

for j in frange(0.1, 1., 0.1):
    nu = j
    executions_time_OCSVM_KDD99CUP = []
    roc_auc_OCSVM_KDD99CUP = []

    #for i in range(5):
############################
    roc_auc_OCSVM_Sub_Datasets = []
    start_OCSVM_KDD99CUP = time()

    # Apply the function on the 6 sub_dataset
    func_OCSVM_KDD99CUP = OneClassSVM(nu=nu)
    func_OCSVM_KDD99CUP.fit(X_sub_dataset_1)
    y_pred_OCSVM_sub_dataset_1 = func_OCSVM_KDD99CUP.predict(X
_sub_dataset_1)

    # Calcul du temps d'exécution
    exec_time_OCSVM_KDD99CUP = time() - start_OCSVM_KDD99CUP
    executions_time_OCSVM_KDD99CUP.append(exec_time_OCSVM_KDD9
9CUP)

    # Calcul de l'aire sous la courbe ROC
    # Apply the function on ROC AUC on the 6 sub_dataset
    auc_OCSVM_KDD99CUP = roc_auc_score(y_transform_sub_dataset
_1, y_pred_OCSVM_sub_dataset_1)
    roc_auc_OCSVM_Sub_Datasets.append(auc_OCSVM_KDD99CUP)

    # Mean of the ROC AUC of the 6 sub_datasets
    roc_auc_OCSVM_KDD99CUP.append(statistics.mean(roc_auc_OCSV
M_Sub_Datasets))
 ########################################

    print("Execution time of OC-SVM applied on KDD99CUP (in se
cond)")
    print(executions_time_OCSVM_KDD99CUP)
    print("ROC AUC for OC-SVM applied on KDD99CUP")
    print(roc_auc_OCSVM_KDD99CUP)
    averages_executions_time_OCSVM_KDD99CUP.append(statistics.
mean(executions_time_OCSVM_KDD99CUP))
    averages_roc_auc_OCSVM_KDD99CUP.append(statistics.mean(roc
_auc_OCSVM_KDD99CUP))
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/ipykernel_launcher.py:4: MatplotlibDeprecationW
arning: numpy.arange
  after removing the cwd from sys.path.
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[2986.6513409614563]
ROC AUC for OC-SVM applied on KDD99CUP
[0.542150334823887]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[3077.2336308956146]
ROC AUC for OC-SVM applied on KDD99CUP
[0.5011811077824525]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[5292.63653588295]
ROC AUC for OC-SVM applied on KDD99CUP
[0.5071569431900786]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[3780.579262971878]
ROC AUC for OC-SVM applied on KDD99CUP
[0.46555457959777708]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)


Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[3394.2336580753326]
ROC AUC for OC-SVM applied on KDD99CUP
[0.4916751932947111]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[3581.1361231803894]
ROC AUC for OC-SVM applied on KDD99CUP
[0.45291208670176447]
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[3182.6861062049866]
ROC AUC for OC-SVM applied on KDD99CUP
[0.4703015067074917]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[2253.939885854721]
ROC AUC for OC-SVM applied on KDD99CUP
[0.5046190276890543]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)

Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[936.0182049274445]
ROC AUC for OC-SVM applied on KDD99CUP
[0.4901677974315483 7]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)
```

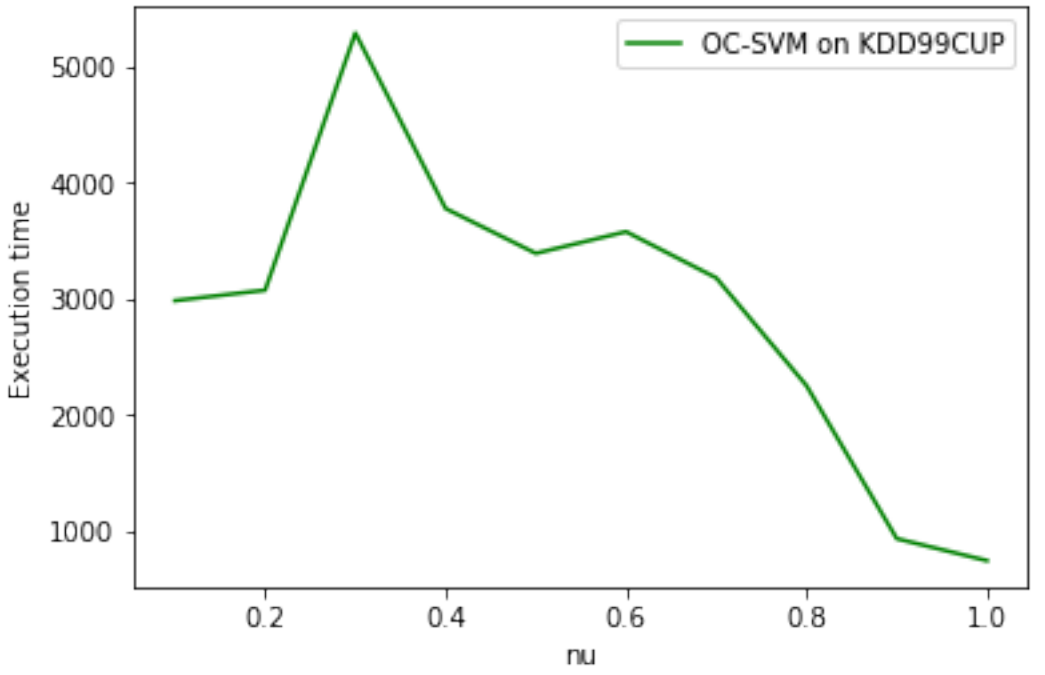
```
Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[746.6968650817871]
ROC AUC for OC-SVM applied on KDD99CUP
[0.5]
```

Résultat de OC-SVM sur KDD99CUP

In [18]:

```
print("Execution time of OC-SVM applied on KDD99CUP (in second
)")
print(averages_executions_time_OCSVM_KDD99CUP)
plt.plot(frange(0.1, 1., 0.1), averages_executions_time_OCSVM_
KDD99CUP,  "g-", label="OC-SVM on KDD99CUP")
#plt.axis([0, 1, 0, 1])
plt.xlabel('nu')
plt.ylabel('Execution time')
plt.legend(loc="best")
plt.show()
```

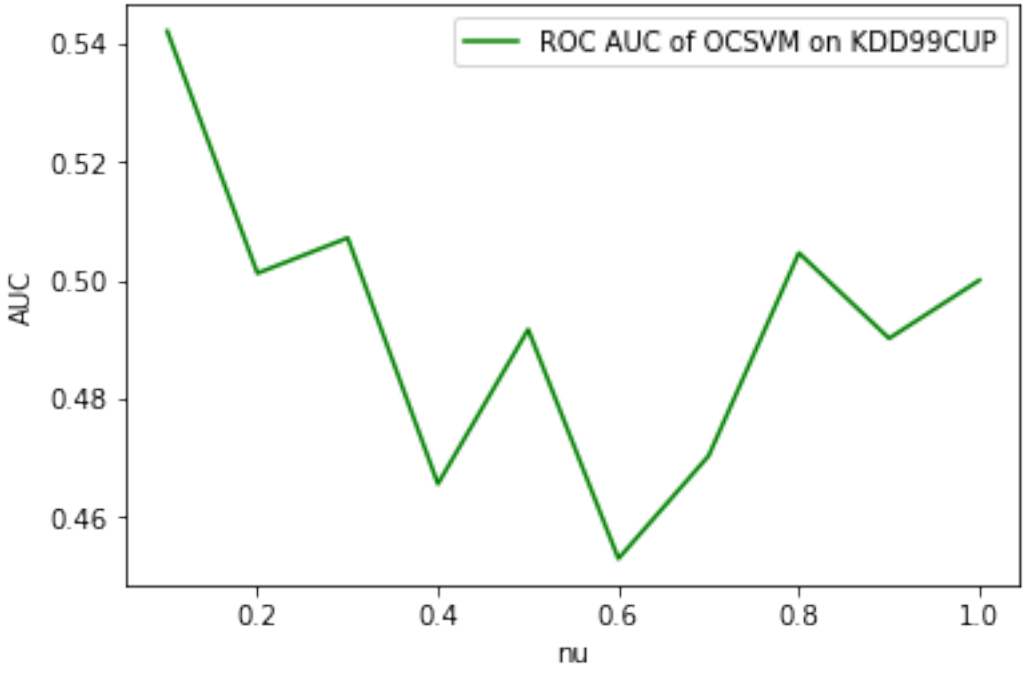
```
Execution time of OC-SVM applied on KDD99CUP (in s
econd)
[2986.6513409614563, 3077.2336308956146, 5292.6365
3588295, 3780.579262971878, 3394.2336580753326, 35
81.1361231803894, 3182.6861062049866, 2253.9398858
54721, 936.0182049274445, 746.6968650817871]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/ipykernel_launcher.py:3: MatplotlibDeprecationW
arning: numpy.arange
  This is separate from the ipykernel package so w
e can avoid doing imports until
```



In [19]:

```
print("ROC AUC for OC-SVM applied on KDD99CUP")
print(averages_roc_auc_OCSVM_KDD99CUP)
plt.plot(frange(0.1, 1., 0.1), averages_roc_auc_OCSVM_KDD99CUP
,  "g-", label="ROC AUC of OCSVM on KDD99CUP")
#plt.axis([0, 1, 0, 1])
plt.xlabel('nu')
plt.ylabel('AUC')
plt.legend(loc="best")
plt.show()
```

```
ROC AUC for OC-SVM applied on KDD99CUP
[0.542150334823887, 0.5011811077824525, 0.50715694
31900786, 0.4655545795977708, 0.4916751932947111,
0.45291208670176447, 0.4703015067074917, 0.5046190
276890543, 0.49016779743154837, 0.5]

/Users/thesard/anaconda3/lib/python3.7/site-packag
es/ipykernel_launcher.py:3: MatplotlibDeprecationW
arning: numpy.arange
  This is separate from the ipykernel package so w
e can avoid doing imports until
```



Interprétation du résultat de OC-SVM sur KDD99CUP

nu= 0.1

Exécution des méthodes avec les meilleurs paramètres trouvés sur KDD99CUP

In [24]:

```
executions_time_KDD99CUP = []
methods = []

# Isolation Forest
methods.append('IForest')
start_time = time()
n_estimators = 280
func_IF_KDD99CUP = IsolationForest(n_estimators=n_estimators)
func_IF_KDD99CUP.fit(X_sub_dataset_1)
y_pred_IF_KDD99CUP = func_IF_KDD99CUP.predict(X_sub_dataset_1)
# Calcul du temps d'exécution
executions_time_KDD99CUP.append(time() - start_time)

# LOF
methods.append('LOF')
start_time = time()
n_neighbors = 195
func_LOF_KDD99CUP = LocalOutlierFactor(n_neighbors=n_neighbors
)
#func_LOF_KDD99CUP.fit(X_sub_dataset_1)
y_pred_LOF_KDD99CUP = func_LOF_KDD99CUP.fit_predict(X_sub_data
set_1)
# Calcul du temps d'exécution
executions_time_KDD99CUP.append(time() - start_time)

# OC-SVM
methods.append('OC-SVM')
start_time = time()
nu = 0.1
func_OCSVM_KDD99CUP = OneClassSVM(nu=nu)
func_OCSVM_KDD99CUP.fit(X_sub_dataset_1)
y_pred_OCSVM_KDD99CUP = func_OCSVM_KDD99CUP.predict(X_sub_data
set_1)
# Calcul du temps d'exécution
executions_time_KDD99CUP.append(time() - start_time)
```

```
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:213: FutureWarning:
default contamination parameter 0.1 will change in
version 0.22 to "auto". This will change the predi
ct method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:223: FutureWarning:
behaviour="old" is deprecated and will be removed
in version 0.22. Please use behaviour="new", which
makes the decision_function change to match other
anomaly detection algorithm API.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/ensemble/iforest.py:417: DeprecationWar
ning: threshold_ attribute is deprecated in 0.20 a
nd will be removed in 0.22.
  " be removed in 0.22.", DeprecationWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/neighbors/lof.py:236: FutureWarning: de
fault contamination parameter 0.1 will change in v
ersion 0.22 to "auto". This will change the predic
t method behavior.
  FutureWarning)
/Users/thesard/anaconda3/lib/python3.7/site-packag
es/sklearn/svm/base.py:196: FutureWarning: The def
ault value of gamma will change from 'auto' to 'sc
ale' in version 0.22 to account better for unscale
d features. Set gamma explicitly to 'auto' or 'sca
le' to avoid this warning.
  "avoid this warning.", FutureWarning)
```

In [25]:

```
precisions_KDD99CUP = []
recalls_KDD99CUP = []
roc_aucs_KDD99CUP = []
f1_scores_KDD99CUP = []

# Isolation Forest
# Precision
precisions_KDD99CUP.append(precision_score(y_transform_sub_dat
aset_1, y_pred_IF_KDD99CUP))
# Recall
recalls_KDD99CUP.append(recall_score(y_transform_sub_dataset_1
, y_pred_IF_KDD99CUP))
# f1_score
f1_scores_KDD99CUP.append(f1_score(y_transform_sub_dataset_1
```

```
f1_scores_KDD99CUP.append(f1_score(y_transform_sub_dataset_1,
y_pred_IF_KDD99CUP))
# Calcul de l'aire sous la courbe ROC
auc_IF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_1, y_p
red_IF_KDD99CUP)
roc_aucs_KDD99CUP.append(auc_IF_KDD99CUP)
print("Confusion Matrice IForest")
print(confusion_matrix(y_transform_sub_dataset_1, y_pred_IF_KD
D99CUP))

# LOF
# Precision
precisions_KDD99CUP.append(precision_score(y_transform_sub_dat
aset_1, y_pred_LOF_KDD99CUP))
# Recall
recalls_KDD99CUP.append(recall_score(y_transform_sub_dataset_1
, y_pred_LOF_KDD99CUP))
# f1_score
f1_scores_KDD99CUP.append(f1_score(y_transform_sub_dataset_1,
y_pred_LOF_KDD99CUP))
# Calcul de l'aire sous la courbe ROC
auc_LOF_KDD99CUP = roc_auc_score(y_transform_sub_dataset_1, y_
pred_LOF_KDD99CUP)
roc_aucs_KDD99CUP.append(auc_LOF_KDD99CUP)
print("Confusion Matrice LOF")
print(confusion_matrix(y_transform_sub_dataset_1, y_pred_LOF_K
DD99CUP))

# OC-SVM
# Precision
precisions_KDD99CUP.append(precision_score(y_transform_sub_dat
aset_1, y_pred_OCSVM_KDD99CUP))
# Recall
recalls_KDD99CUP.append(recall_score(y_transform_sub_dataset_1
, y_pred_OCSVM_KDD99CUP))
# f1_score
f1_scores_KDD99CUP.append(f1_score(y_transform_sub_dataset_1,
y_pred_OCSVM_KDD99CUP))
# Calcul de l'aire sous la courbe ROC
auc_OCSVM_KDD99CUP = roc_auc_score(y_transform_sub_dataset_1,
y_pred_OCSVM_KDD99CUP)
roc_aucs_KDD99CUP.append(auc_OCSVM_KDD99CUP)
print("Confusion Matrice OC-SVM")
print(confusion_matrix(y_transform_sub_dataset_1, y_pred_OCSVM
_KDD99CUP))
```

```
Confusion Matrice IForest
[[  176     0]
 [10159 93016]]
Confusion Matrice LOF
[[  119    57]
 [10216 92959]]
Confusion Matrice OC-SVM
[[   89    87]
 [43476 59699]]
```

Présentation des résultats sur KDD99CUP

Résultat : Temps d'exécution, précision, rappel, score f1, AUC

In [26]:

```
print("Methods ===>")
print(methods)
print("Execution Time ===>")
print(executions_time_KDD99CUP)
print("Precision ===>")
print(precisions_KDD99CUP)
print("Recall ===>")
print(recalls_KDD99CUP)
print("f1_score ===>")
print(f1_scores_KDD99CUP)
print("ROC AUC ===>")
print(roc_aucs_KDD99CUP)
```

```
Methods ===>
['IForest', 'LOF', 'OC-SVM']
Execution Time ===>
[28.918054819107056, 35.97923183441162, 4615.74405
4079056]
Precision ===>
[1.0, 0.9993872022017717, 0.9985448098216974]
Recall ===>
[0.9015362248606736, 0.9009837654470559, 0.5786188
514659559]
f1_score ===>
[0.9482188275710914, 0.9476377611613173, 0.7326783
708985586]
ROC AUC ===>
[0.9507681124303369, 0.7885600645417098, 0.5421503
34823887]
```

Interprétation des résultats en fonction du Temps d'exécution, précision, rappel, score f1, AUC sur KDD99CUP

IsolationForest a eu de meilleures performances que les autres. Entre OC-SVM et LOF : OC-SVM a découvert et bien classé plus d'anomalies que LOF. Par contre, LOF a découvert et bien placé plus d'observations normales que OC-SVM. La question est donc de savoir ce qui importe le plus entre la découverte des anomalies et celle des observations normales car c'est en fonction de l'importance que nous donnons à l'un ou à l'autre que nous pourront dire la méthode la meilleure entre OC-SVM et LOF. Cette question fait remarquer qu'en réalité, toutes les méthodes ont sont performantes dans des conditions données en fonction de ce que l'on souhaite obtenir comme résultat. Dans ce cas précis, OC-SVM lance

beaucoup de fausse alertes mais attrappe beaucoup d'anomalies. Par contre, LOF n'attrappe pas beaucoup d'anomalies mais reconnais beaucoup de données normales. On peut peut-être se dire que dans le cas espèce OC-SVM est mieux dans la mesure où une méthode de détection d'anomalies a pour but principal de détecter les anomalies et donc laisser passer des anomalies en ferait une méthode moins bonne et que les observations normales que OC-SVM aurait classé comme anormales sont peut être des nouveautés. Considérant l'aire sous la courbe ROC, nous avons l'ordre suivant IForest - OC-SVM et LOF

Confusion Matrice IForest [[176 0] [10159 93016]] Confusion Matrice LOF [[119 57] [10216 92959]] Confusion Matrice OC-SVM [[89 87] [43476 59699]]

Tableau du résultat TOBC = VN/(VN + FP) = Spécifité TNBC = VP/(FN + VP)

Méthode |Précision | Rappel | Score f1 | AUC | TOBC | TNBC | CPU Time |

IForest |1.0 |0.9015362|0.9482188 |0.9507681|1. |0.9015362|28.9180548 |

OC-SVM |0.9985448 |0.5786188|0.7326783 |0.5421503|0.5056818|0.5786189|4615.7440540|

LOF |0.9993872 |0.9009837|0.9476377 |0.7885600|0,6761364|0.9009838|35.9792318 |

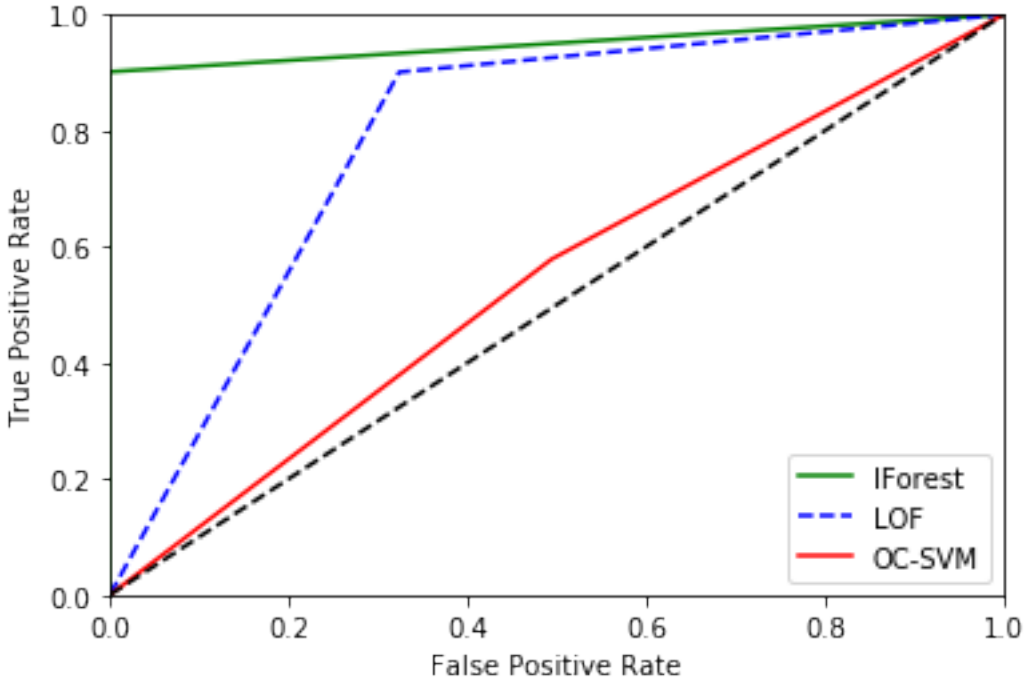
Resultat: Courbe ROC pour KDD99CUP

In [29]:

```
fpr_IF_KDD99CUP, tpr_IF_KDD99CUP, thresholds_IF_KDD99CUP = roc
_curve(y_transform_sub_dataset_1, y_pred_IF_KDD99CUP)
fpr_LOF_KDD99CUP, tpr_LOF_KDD99CUP, thresholds_LOF_KDD99CUP =
roc_curve(y_transform_sub_dataset_1, y_pred_LOF_KDD99CUP)
fpr_OCSVM_KDD99CUP, tpr_OCSVM_KDD99CUP, thresholds_OCSVM_KDD99
CUP = roc_curve(y_transform_sub_dataset_1, y_pred_OCSVM_KDD99C
UP)
```

In [30]:

```
plt.plot(fpr_IF_KDD99CUP, tpr_IF_KDD99CUP, "g-", label="IFores
t")
plt.plot(fpr_LOF_KDD99CUP, tpr_LOF_KDD99CUP, "b--", label="LOF
")
plt.plot(fpr_OCSVM_KDD99CUP, tpr_OCSVM_KDD99CUP, "r-", label="
OC-SVM")
plt.plot([0, 1], [0, 1], 'k--')
plt.axis([0, 1, 0, 1])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.legend(loc="lower right")

plt.show()
```



Interprétation de la courbe ROC

La courbe ROC mesurant le taux de faux positif par rapport au taux de vrai positif, plus la courbe s'éloigne de celle oblique en noir, mieux c'est. Ainsi, on peut dire que pour KDD99CUP, IForest est la meilleure méthode de détection d'anomalies suivi par OC-SVM. Le classement de LOF pourrait s'expliquer par le fait que les anomalies qui y sont présentes sont des anomalies global alors que LOF est doué pour les anomalies locales. POur vérifier ça peut être qu'on peut utiliser une méthode de clustering comme K-means pour voir comment les anomalies seront classées par cette méthode.

Conclusion KDD99CUP

Pour un petit jeu de données statique de faible dimension, Isolation forest a tendance à mieux performer et semble être plus rapide que OC-SVM et LOF. OC-SVM détecte beaucoup d'anomalies avec un fort taux de fausses alertes et un temps d'exécution plus grand que LOF. Par contre, LOF détecte moins d'anomalies avec un faible taux de fausses alertes. Notons que la performance des méthodes dépends non seulement du jeu de données considéré mais aussi des valeurs des hypermaramètres. Pour LOF et IsolationForest, le temps d'exécution semble être en correlation positive avec respectivement le nombre de voisins et nombre de d'estimateurs. Par contre, le temps d'exécution serait en correlation négative avec nu.