

# MO444/MC886 - Aprendizado de máquina

## Exercício 02

José Ernesto Stelzer Monar - 139553

5 de outubro de 2016

### Código da solução

```
import numpy as np
from sklearn.cross_validation import KFold, StratifiedKFold
from sklearn.svm import SVC

np.set_printoptions(threshold=np.inf)

realData = []
classification = []

cVals = [2**-5, 2**-2, 2**0, 2**2, 2**5]
gammaVals = [2**-15, 2**-10, 2**-5, 2**0, 2**5]

#Leitura e processamento dos dados
f = open('data1.csv', 'r')
data = f.readlines()
f.close()
for i in range(1, len(data)):
    data[i] = data[i].split(',')
    realData.append([float(d) for d in data[i][::-1]])
    classification.append(int(data[i][-1].strip()))
x = np.array(realData)
y = np.array(classification)

skf = StratifiedKFold(y, n_folds=5)

meanAcc = 0
i = 1
for outTrainIndex, outTestIndex in skf:
    bAcc = 0

    # Dados considerando os folds de treino
    xIn = x[outTrainIndex]
    yIn = y[outTrainIndex]
```

```

for c in cVals:
    for gamma in gammaVals:
        kf = KFold(len(outTrainIndex), n_folds=3)
        # Média das acurácias para a combinação (c,gamma) dos 3 folds
        acc = 0
        for inTrainIndex, inTestIndex in kf:
            svc = SVC(C=c, gamma=gamma, kernel='rbf')
            svc.fit(xIn[inTrainIndex], yIn[inTrainIndex])
            acc += svc.score(xIn[inTestIndex], yIn[inTestIndex])
        acc /= 3
        # Escolhe a melhor combinação de hiperparametros baseada na média das acurá
        if (acc > bAcc):
            bC = c
            bGamma = gamma
            bAcc = acc
    svc = SVC(C=bC, gamma=bGamma, kernel='rbf')
    svc.fit(x[outTrainIndex], y[outTrainIndex])
    acc = svc.score(x[outTestIndex], y[outTestIndex])
    # print("Fold ", i, ":")
    # print(2*" ", "c = ", bC)
    # print(2*" ", "gamma = ", bGamma)
    # print(2*" ", "accuracy = ", acc)
    meanAcc += acc
    i += 1

print("Acurácia média da validação de fora: ", meanAcc/5)

bAcc = 0
for c in cVals:
    for gamma in gammaVals:
        acc = 0
        kf = KFold(len(y), n_folds=3)
        for inTrainIndex, inTestIndex in kf:
            svc = SVC(C=c, gamma=gamma, kernel='rbf')
            svc.fit(x[inTrainIndex], y[inTrainIndex])
            acc += svc.score(x[inTestIndex], y[inTestIndex])
        acc /= 3
        if (acc > bAcc):
            bC = c
            bGamma = gamma
            bAcc = acc
print("\nHiperparâmetros do classificador final:")
print(2*" ", "c = ", bC)
print(2*" ", "gamma = ", bGamma)

```

## Saída da execução

Acurácia média da validação de fora: 0.911927025009

Hiperparâmetros do classificador final:

c = 4

gamma = 0.03125

## Respostas

- 1 Acurácia média da validação de fora foi 0.911927025009
- 2 Os valores dos hiperparâmetros escolhidos para o classificador final foram:  $c = 4$  e  $\gamma = 0.03125$