**《机器学习与数据挖掘》实验五**

实验题目： 用线性核与高斯核训练支持向量机

实验目的： 掌握支持向量机的原理及应用

实验环境（硬件和软件） Anaconda/Jupyter notebook/Pycharm

实验内容：

使用Sklearn，在西瓜集3.0上分别使用线性核和高斯核训练一个SVM，并比较其支持向量的差别。

**要求：**

**一、**已经给定部分代码，补充完整的代码，需要补充代码的地方已经用红色字体标注，包括：

**（1）#补充构建SVM模型及训练代码**

**（2）#补充预测代码**

**二、**将补充完整的代码提交，并提交实验结果；（**也可以自己重写这部分的代码提交**）

data\_file\_watermelon\_3a = "watermelon\_3a.csv"

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

df = pd.read\_csv(data\_file\_watermelon\_3a, header=None, )

df.columns = ['id', 'density', 'sugar\_content', 'label']

df.set\_index(['id'])

X = df[['density', 'sugar\_content']].values

y = df['label'].values

########## SVM training and comparison

# based on linear kernel as well as gaussian kernel

from sklearn import svm

for fig\_num, kernel in enumerate(('linear', 'rbf')):

**#补充构建SVM模型及训练代码**

#给定新的样本X\_test，预测其标签

X\_test = [[0.719,0.103]]

**#补充预测代码**

# get support vectors

sv = svc.support\_vectors\_

##### draw decision zone

plt.figure(fig\_num)

plt.clf()

# plot point and mark out support vectors

plt.scatter( X[:,0], X[:,1], edgecolors='k', c=y, cmap=plt.cm.Paired, zorder=10)

plt.scatter(sv[:,0], sv[:,1], edgecolors='k', facecolors='none', s=80, linewidths=2, zorder=10)

# plot the decision boundary and decision zone into a color plot

x\_min, x\_max = X[:, 0].min() - 0.2, X[:, 0].max() + 0.2

y\_min, y\_max = X[:, 1].min() - 0.2, X[:, 1].max() + 0.2

XX, YY = np.meshgrid(np.arange(x\_min, x\_max, 0.02), np.arange(y\_min, y\_max, 0.02))

Z = svc.decision\_function(np.c\_[XX.ravel(), YY.ravel()])

Z = Z.reshape(XX.shape)

plt.pcolormesh(XX, YY, Z>0, cmap=plt.cm.Paired)

plt.contour(XX, YY, Z, colors=['k', 'k', 'k'], linestyles=['--', '-', '--'], levels=[-.5, 0, .5])

plt.title(kernel)

plt.axis('tight')

plt.show()