**作业6.8**

1. **实验要求**

以西瓜数据集3.0a的密度为输入，含糖率为输出，试使用LIBSVM训练一个SVR。

1. **实验原理**

主要是通过升维后，在高维空间中构造线性决策函数来实现线性回归，用e不敏感函数时，其基础主要是 e 不敏感函数和核函数算法。

若将拟合的数学模型表达多维空间的某一曲线，则根据e 不敏感函数所得的结果，就是包括该曲线和训练点的“ e管道”。在所有样本点中，只有分布在“管壁”上的那一部分样本点决定管道的位置。这一部分训练样本称为“支持向量”。为适应训练样本集的非线性，传统的拟合方法通常是在线性方程后面加高阶项。此法诚然有效，但由此增加的可调参数未免增加了过拟合的风险。支持向量回归算法采用核函数解决这一矛盾。用核函数代替线性方程中的线性项可以使原来的线性算法“非线性化”，即能做非线性回归。与此同时，引进核函数达到了“升维”的目的，而增加的可调参数是过拟合依然能控制。

1. **实验过程与代码**

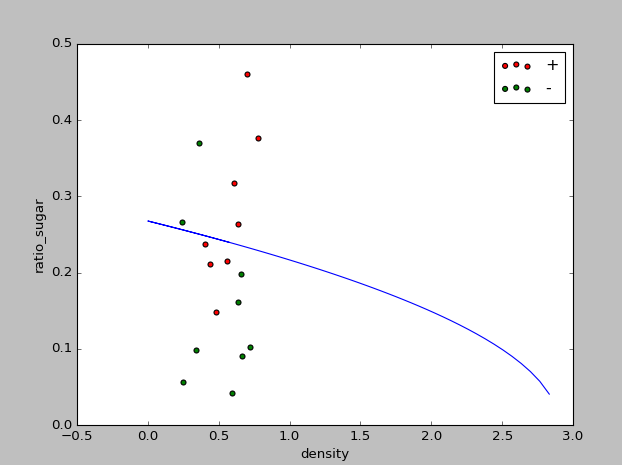
与SVM类似，利用sklearn的svr函数进行训练，过程如下：

1. 准备训练集与标记。
2. 利用sklearn的svr函数进行拟合得到相关参数。
3. 得到a,w,b等值并代入到自己给定的横坐标中得到线性方程。
4. 画出图像。

代码如下：

**from** sklearn.svm **import** SVR  
**from** numpy **import** \*  
**import** matplotlib.pyplot **as** plt  
*#data introduce*data = mat([[0.697,0.460,1],  
 [0.774,0.376,1],  
 [0.634,0.264,1],  
 [0.608,0.318,1],  
 [0.556,0.215,1],  
 [0.403,0.237,1],  
 [0.481,0.149,1],  
 [0.437,0.211,1],  
 [0.666,0.091,0],  
 [0.243,0.267,0],  
 [0.245,0.057,0],  
 [0.343,0.099,0],  
 [0.639,0.161,0],  
 [0.657,0.198,0],  
 [0.360,0.370,0],  
 [0.593,0.042,0],  
 [0.719,0.103,0]])  
x=data[:,0:2]  
y=data[:,2]  
clf=SVR(kernel=**'linear'**,C=1000)  
clf.fit(x,y)  
x\_min, x\_max = x[:, 0].min() - 1, x[:, 0].max() + 1  
y\_min, y\_max = x[:, 1].min() - 1, x[:, 1].max() + 1  
xx=arange(x\_min, x\_max, 0.02)  
w = clf.coef\_[0]  
*#print(w)*t=square(xx)  
a = -w[0] / w[1]  
y1 =sqrt( a \* t- clf.intercept\_[0] / w[1])  
X0=data[0:8,:]  
X1=data[8:17,:]  
plt.scatter(X0[:,0],X0[:,1],c=**'r'**,label=**'+'**)  
plt.scatter(X1[:,0],X1[:,1],c=**'g'**,label=**'-'**)  
plt.plot(t,y1)  
plt.xlabel(**'density'**)  
plt.ylabel(**'ratio\_sugar'**)  
plt.legend()  
plt.show()

**4、实验结果如下：**



可以将两者分开的。