1. **SVM<-function(traindata,trainrowname = NULL, traincolname = NULL, testdata = NULL, testrowname = NULL, testcolname = NULL, yname=NULL, xname=NULL, scale = TRUE, costrange = NULL, gammarange = NULL, sampling = "fix", plotstr = plotstr, svmname = svmname)**

功能：

支持向量机。，需要加载e1071包。本脚决策树采用高斯核函数算法，并采用网格搜索寻找最优参数，参数的范围由costrange和gammarange给定。

输入：

**traindata:**。训练数据集合，两列及以上。

其中yname指定的列是类标号，类型为字符串向量，数值型向量 会被自动转为字符串型。

其中xname定的列是自变量，全部为数值型向量；字符串型数字向量(‘1’,’2’,’3’)会被自动转为数值型（1，2，3）; 字符串型非数字向量（“男”，“女”）会被自动转为（1，2），但是强烈建议事先转化字符串型非数字向量。自变量个数不宜过少，一般至少2个。

**testdata** ：测试数据集合，当模型训练完后，输入测试集数据，用来预测其类标号。要求testdata的列名和类型要与xname追定的训练集保持一致。例如，traindata 中xname = [‘’age”, “sex”], 则testdata列也要是[‘’age”, “sex”]这两个向量，且数据类型保持一致。如果testdata为空，则不做预测。

**trainrowname**：训练集行名。

**traincolname**：训练集列名。

**testrowname**：测试集行名。

**testcolname**：测试集列名。

**yname**：因变量名称，类型为字符串，不可缺省。

**xname**：自变量名称，类型为字符串向量，不可缺省。

**scale:** 是否标准化dataset数据（减去均值，除以标准差），缺省标准化。

**costrange**：正则化参数取值范围，数值型向量，只能包含两个数值。例如，**costrange = [-1,1]**表示模型中正则化参数的取值范围是2^（-1:1） = [2^(-1), 2^0, 2^1]。

**gammarange**：高斯核函数参数取值范围，数值型向量，只能包含两个数值。例如，**gammarange = [-1,1]**表示模型中高斯核函数参数的取值范围是2^（-1:1） = [2^(-1), 2^0, 2^1]

**sampling:** 模型验证方式，用去确定最优参数， 类型为字符串。

“fix”表示把traindata随机分为训练集和验证集。

‘cross’表示实用10折交叉验证。

“boot” 表示重抽样。

**plotstr**：图片输出文件夹目录，类型为字符串，缺省不输出。

**svmname**：参数选择等高线图输出名称，类型为字符串，缺省不输出。

输出：

**BestCost**：给定搜索范围下的最优正则化参数

**BestGamma**：给定搜索范围下的最优高斯核函数参数

**BestTrainError**：训练集上最优的训练误差。

**ConfusionRowName**：训练集上的分类混淆矩阵的行名，类型为字符串向量

**ConfusionColName**：训练集上的分类混淆矩阵的列名，类型为字符串向量

**Confusion**：训练集上的分类混淆矩阵。

如果testdata不为空，则输出预测结果：

**PredictionRowName**：预测结果的的行名（即testrowname），类型为字符串向量，。

**PredictionColName**：预测结果的列名，类型为字符串向量。

**pred**：预测结果。

注：svm的复杂度和样本量有关。大样本下会比较耗时。**costrange**和

**gammarange**范围越大也越耗时。**cross** 和**boot** 的计算量大约是**fix**的**10**

倍。