**Ici数据部分**

结果分为两部分，文件名分别为ici\_result\_classification.csv，ici\_result\_regression.csv，是分类和回归分析的结果。

**分类实验：**

linear/rbf是核函数的类型，每一种核函数设定了三个c值，每个c值跑了两次实验：single是用单个feature运行出的结果，reverse是用除去这个feature以外的所有feature运行出的结果。所以结果共2\*3\*2=12列。

从单分类结果来看，feature V的结果最好，举例为52.035（linear/c=1/single）；从其余分类结果来看，除去feature K的结果最差，举例为30.7058（linear/c=1/reverse），即feature K对正确分类的影响最大。

**回归实验：**

epsilon-SVR/nu-SVR是核函数的类型，每个核函数也分为single与reverse实验，每次实验有两列结果：MSR = Mean squared error，越小越好；SCC = Squared correlation coefficient，约接近1越好。结果共2\*2\*2=8列。

从单分类结果来看，feature K的结果最好，MSR=4.14202,SCC=0.216261（epsilon-SVR/single）；从其余分类结果来看，则各个feature之间没有明显差别。

综上，ici数据部分的feature K对数据影响最大。

**工厂数据部分**

结果分为两部分，文件名分别为factory\_result\_classification.csv，factory\_result\_regression.csv，是分类和回归分析的结果。

工厂数据主要的处理是将所有文字型数据变为0，忽略所有起止时间的feature（保留持续时间的feature），数值为NA的部分变为-1，合格产品结果为1，忽略不合格产品的错误类型而把所有不合格产品的结果全部设为-1，去掉所有的无效数据等等。

结果数据的表头，数据举例的类型与ici的表格相同。

**分类实验：**

在reverse结果区别不大的情况下，feature AG的单分类single实验的正确率达到了100%。

**回归实验：**

feature AG的MSR=0.0110247是最低的方差值，其SCC=0.0003。

综上，工厂数据部分的feature AG对数据影响最大。具体影响方式是feature AG为10时产品合格，为11时不合格。

注：鉴于feature AG的title为LAST\_RESULT，可能feature AG只是结果的另外一种表述，不是影响产品是否合格的特征。

**R语言**

data <- read.csv("ici\_trans.csv", header=FALSE)

inputData <- data.frame(data[, c (11,14,22)], response = as.factor(data[, c(28)]))

svmfit <- svm(response ~ ., data = inputData, kernel = "radial", cost = 10, scale = FALSE)

mean(inputData$response == predict(svmfit))

data <- read.csv("factory\_trans.csv", header=FALSE)

inputData <- data.frame(data[, c (33)], response = as.factor(data[, c(25)]))

svmfit <- svm(response ~ ., data = inputData, kernel = "radial", cost = 10, scale = FALSE)

mean(inputData$response == predict(svmfit))