In [ ]:

```
# R语言

# Author :坚定的唯物主义鼠鼠

#R BCancer 随机森林
#数据来源https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

library(randomForest)
library(caret)
set.seed(1234)
```

我们读取数据，并且取数据中的前519行作为训练集，后50行作为测试集

In [ ]:

```
data=read.csv("./data/data_处理后.csv")
#这里的处理指的是将数据中的B、M转换为0、1

tdata=data[-1]
#取tdata的前519行
train=tdata[1:519,]
#取tdata的到最后部分
test=tdata[520:569,]
```

In [ ]:

```
rf=randomForest(as.factor(diagnosis)~.,data=train,na.action=na.roughfix,importar
rf
```

```
Call:
 randomForest(formula = as.factor(diagnosis) ~ ., data = train,      importance
= TRUE, ntree = 5000, na.action = na.roughfix)
               Type of random forest: classification
                     Number of trees: 5000
No. of variables tried at each split: 5

        OOB estimate of  error rate: 3.66%
Confusion matrix:
    0   1 class.error
0 309   8  0.02523659
1  11 191  0.05445545
```

In [ ]:

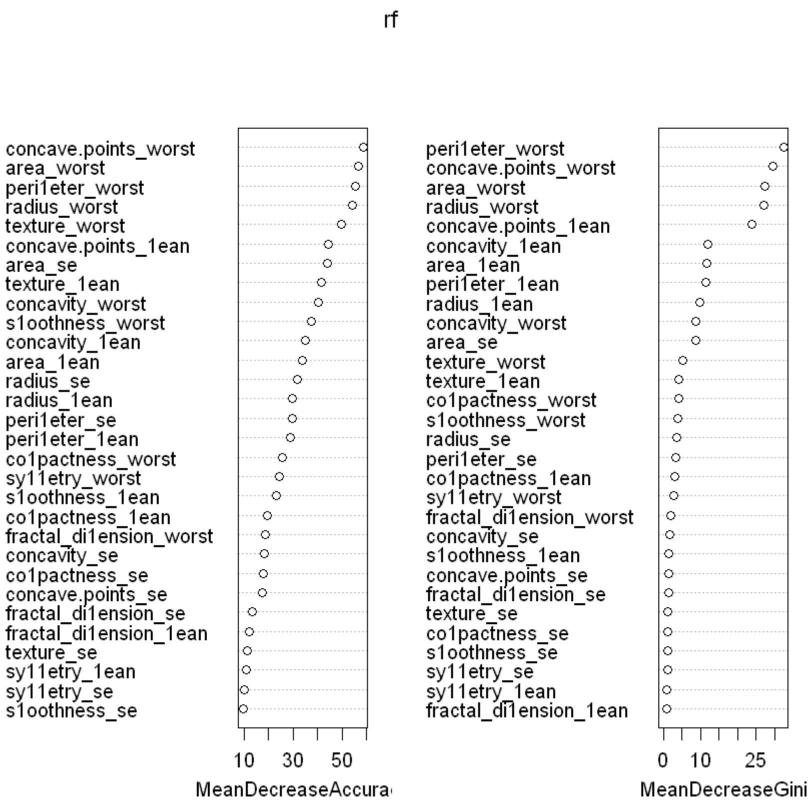
```
out=importance(rf)  #计算变量重要性(对结果影响的权重)
out
```

A matrix: 30 × 4 of type dbl

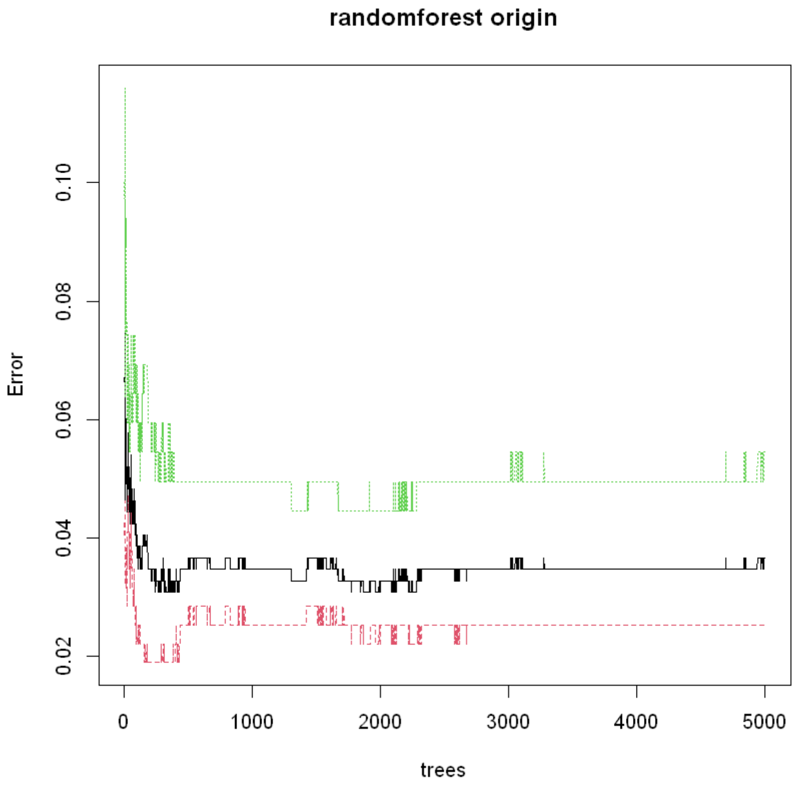
	0	1	MeanDecreaseAccuracy	MeanDecreaseGini
radius_1ean	26.112147	17.809744	29.714210	9.8866217
texture_1ean	30.063144	33.000697	41.350285	4.2467905
peri1eter_1ean	23.953476	18.066851	29.038723	11.3981788
area_1ean	30.219095	18.513597	33.571503	11.6018661
s1oothness_1ean	8.106796	21.661381	23.174832	1.5262526
co1pactness_1ean	14.568648	13.050985	19.375427	3.0504222
concavity_1ean	21.957396	27.216753	35.111412	11.8611796
concave.points_1ean	29.425419	34.441293	44.264160	23.8602257
sy11etry_1ean	3.998218	10.711723	10.873887	0.9954036
fractal_di1ension_1ean	10.104765	5.628074	12.051859	0.8543356
radius_se	26.257203	18.269241	31.789633	3.5150348
texture_se	8.482980	7.380432	11.328754	1.1316575
peri1eter_se	22.689242	19.001134	29.683601	3.3015735
area_se	36.184434	25.279556	44.074501	8.6683929
s1oothness_se	8.435562	3.925434	9.590131	1.0563676
co1pactness_se	15.071959	7.686829	17.712989	1.1228609
concavity_se	12.852565	13.260094	18.229645	1.6382556
concave.points_se	15.370816	8.717615	17.561372	1.3190484
sy11etry_se	9.298527	3.777097	10.111868	1.0199066
fractal_di1ension_se	13.480288	4.157929	13.264384	1.3042096
radius_worst	45.606291	35.767503	54.108122	27.2532416
texture_worst	38.402181	38.440527	49.576111	5.0764932
peri1eter_worst	44.872233	38.632682	55.449010	32.4521630
area_worst	46.552770	39.779338	56.659358	27.4638906
s1oothness_worst	25.280226	29.625672	37.452021	3.7335295
co1pactness_worst	18.261866	19.758629	25.630635	4.1803436
concavity_worst	22.330076	33.347776	40.399577	8.6687363
concave.points_worst	45.911594	38.560623	58.533147	29.4660825
sy11etry_worst	16.410160	20.511613	24.370134	2.7766774
fractal_di1ension_worst	13.825791	12.250829	18.634334	1.8956551

In [ ]:

```
varImpPlot(rf) #画出随机森林中不同变量的重要性,
```



```
In [ ]: plot(rf,main="randomforest origin")#画出随机森林的重要性
```



在Reference和Prediction中，我们的预测值和实际值完全相同，这说明我们的模型是正确的，随机森林模型的准确率是100% 。

我们此次的预测结果是正确的，P-Value的值为1.427e-05，这个值小于0.05，所以我们可以认为我们的模型是正确的。

In [ ]:

```
forest.pred=predict(rf,test,type="class")
forest.cf=confusionMatrix(as.factor(forest.pred),as.factor(test$diagnosis))
forest.cf
```

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 40  0
         1  0 10

               Accuracy : 1
                 95% CI : (0.9289, 1)
    No Information Rate : 0.8
    P-Value [Acc > NIR] : 1.427e-05

                  Kappa : 1

 Mcnemar's Test P-Value : NA

            Sensitivity : 1.0
            Specificity : 1.0
         Pos Pred Value : 1.0
         Neg Pred Value : 1.0
             Prevalence : 0.8
         Detection Rate : 0.8
   Detection Prevalence : 0.8
      Balanced Accuracy : 1.0

       'Positive' Class : 0
```