

1 Scheduling Data

These data consist of information on 4331 jobs in a high performance computing environment. Seven attributes were recorded for each job along with a discrete class describing the execution time.

The predictors are:

Protocol (the type of computation),

Compounds (the number of data points for each jobs),

InputFields (the number of characteristic being estimated),

Iterations (maximum number of iterations for the computations),

NumPending (the number of other jobs pending at the time of launch),

Hour (decimal hour of day for launch time), [Day] (of launch time).

The classes are: VF (very fast), F (fast), M (moderate) and L (long).

```
library(AppliedPredictiveModeling)
data(schedulingData)

library(caret)
set.seed(733)
inTrain <- createDataPartition(schedulingData$Class, p = .75,
  list = FALSE)

training <- schedulingData[ inTrain,]
testing <- schedulingData[-inTrain,]
```

```
> dim(schedulingData)
[1] 4331    8
>
> dim(training)
[1] 3251    8
>
> dim(testing)
[1] 1080    8
>
```

```
library(C50)
oneTree <- C5.0(Class ~ ., data = training)
```

```
> oneTree
```

Call:

```
C5.0.formula(formula = Class ~ ., data = training)
```

Classification Tree

Number of samples: 3251

Number of predictors: 7

Tree size: 199

Non-standard options: attempt to group attributes

```
oneTreePred <- predict(oneTree, testing)
oneTreeProbs <- predict(oneTree, testing, type = "prob")
postResample(oneTreePred, testing$Class)
```

```
> table(testing$Class, oneTreePred)
```

	oneTreePred			
	VF	F	M	L
VF	512	38	2	0
F	50	256	26	4
M	6	46	67	9
L	0	7	10	47