

1 Sonar Data Set

The Sonar data consist of 208 data points collected on 60 predictors. The goal is to predict the two classes The Sonar data consist of 208 data points collected on 60 predictors. The goal is to predict the two classes (M for metal cylinder or R for rock).

M Metal cylinder

R Rock

First, we split the data into two groups: a training set and a test set. To do this, the `createDataPartition` function is used

```
library(caret)
library(mlbench)
data(Sonar)
set.seed(107)
```

```
> head(Sonar)
      V1      V2      V3      V4      V5      V6      V7      V8
1 0.0200 0.0371 0.0428 0.0207 0.0954 0.0986 0.1539 0.1601
2 0.0453 0.0523 0.0843 0.0689 0.1183 0.2583 0.2156 0.3481
3 0.0262 0.0582 0.1099 0.1083 0.0974 0.2280 0.2431 0.3771
4 0.0100 0.0171 0.0623 0.0205 0.0205 0.0368 0.1098 0.1276
5 0.0762 0.0666 0.0481 0.0394 0.0590 0.0649 0.1209 0.2467
6 0.0286 0.0453 0.0277 0.0174 0.0384 0.0990 0.1201 0.1833
```

1.1 Partial least squares

Partial least squares regression (PLS regression) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of minimum variance between

the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the X and Y data are projected to new spaces, the PLS family of methods are known as bilinear factor models. Partial least squares Discriminant Analysis (PLS-DA) is a variant used when the Y is categorical.

```
training <- Sonar[ inTrain,]
testing <- Sonar[-inTrain,]
nrow(training)
nrow(testing)

plsFit <- train(Class ~ .,
  data = training,
  method = "pls",
  ## Center and scale the predictors for the training
  ## set and all future samples.
  preProc = c("center", "scale"))
```

```
ctrl <- trainControl(method = "repeatedcv", repeats = 3)
plsFit <- train(Class ~ .,
data = training,
method = "pls",
tuneLength = 15,
trControl = ctrl,
preProc = c("center", "scale"))
```