Model Performance - Multivariate Adaptive Regression Splines

Sanatan Das

April 24, 2018

Contents

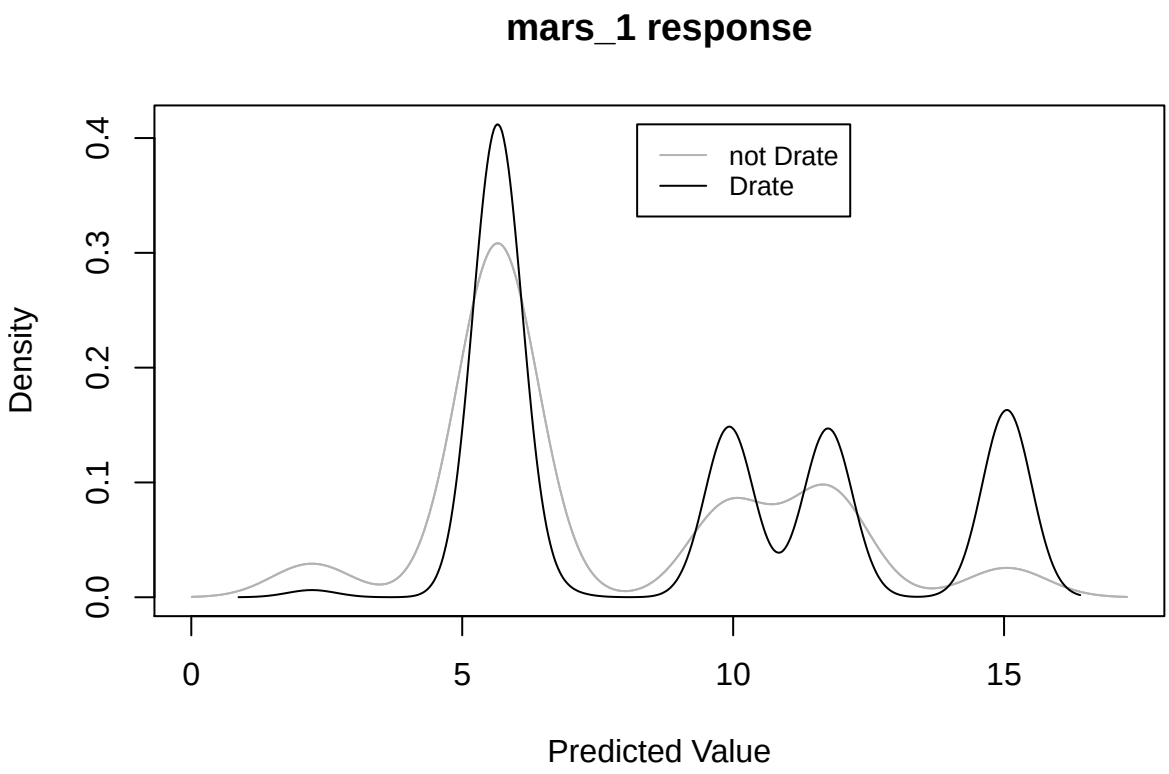
Basic Models (MARS)

Model evaluation on 'ProgLength' feature

```
# MARS model on 'ProgLength' feature
mars_1 <- earth(Drate ~ ProgLength, data = default_rates, pmethod="backward", nprune=20, nfold=10);
summary(mars_1, digit=3);
```

```
## Call: earth(formula=Drate~ProgLength, data=default_rates,
##             pmethod="backward", nprune=20, nfold=10)
##
##              coefficients
## (Intercept)          9.93
## ProgLength11        -3.24
## ProgLength3          1.82
## ProgLength5          5.12
## ProgLength7         -7.70
## ProgLength8         -4.28
##
## Selected 6 of 6 terms, and 5 of 10 predictors
## Termination condition: RSq changed by less than 0.001 at 6 terms
## Importance: ProgLength8, ProgLength5, ProgLength7, ProgLength3, ...
## Number of terms at each degree of interaction: 1 5 (additive model)
## GCV 35.6  RSS 817205  GRSq 0.277  RSq 0.278  CVRSq 0.277
##
## Note: the cross-validation sd's below are standard deviations across folds
##
## Cross validation:   nterms 6.00 sd 0.00    nvars 5.00 sd 0.00
##
##      CVRSq    sd     MaxErr   sd
##      0.277 0.013       50.8 6.12
```

```
plotd(mars_1)
```



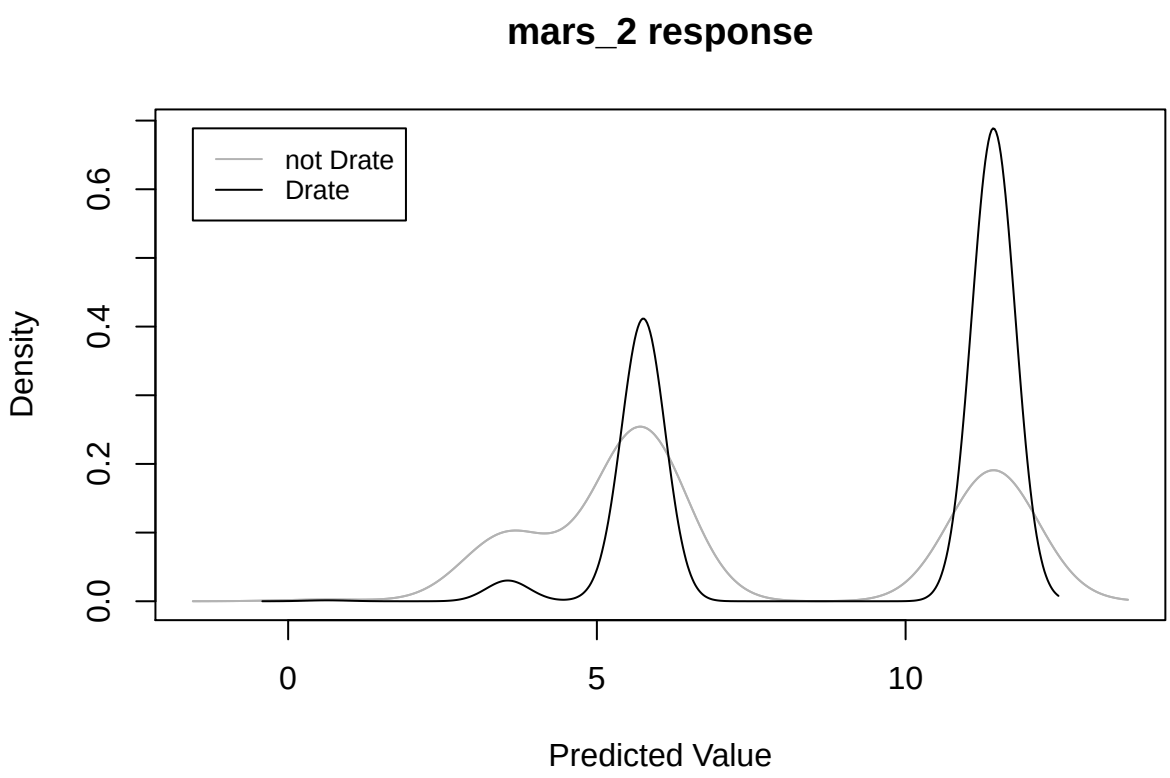
Model evaluation on 'SchoolType' feature

```
# MARS model on 'SchoolType' feature
mars_2 <- earth(Drate ~ SchoolType, data = default_rates, pmethod="backward", nprune=20, nfold=10);
summary(mars_2, digit=3);
```

```
## Call: earth(formula=Drate~SchoolType, data=default_rates,
##             pmethod="backward", nprune=20, nfold=10)
##
##             coefficients
## (Intercept)        11.42
## SchoolType2        -5.67
## SchoolType5        -7.87
## SchoolType6        -6.19
## SchoolType7       -10.79
##
## Selected 5 of 5 terms, and 4 of 5 predictors
## Termination condition: RSq changed by less than 0.001 at 5 terms
## Importance: SchoolType2, SchoolType5, SchoolType6, SchoolType7, ...
## Number of terms at each degree of interaction: 1 4 (additive model)
## GCV 40.8  RSS 937185  GRSq 0.171  RSq 0.172  CVRSq 0.172
##
## Note: the cross-validation sd's below are standard deviations across folds
##
## Cross validation:   nterms 5.00 sd 0.00    nvars 4.00 sd 0.00
```

```
##
##      CVRSq   sd     MaxErr   sd
##      0.172 0.01       51.1 5.62
```

```
plotd(mars_2)
```



Model evaluation on 'Num' feature

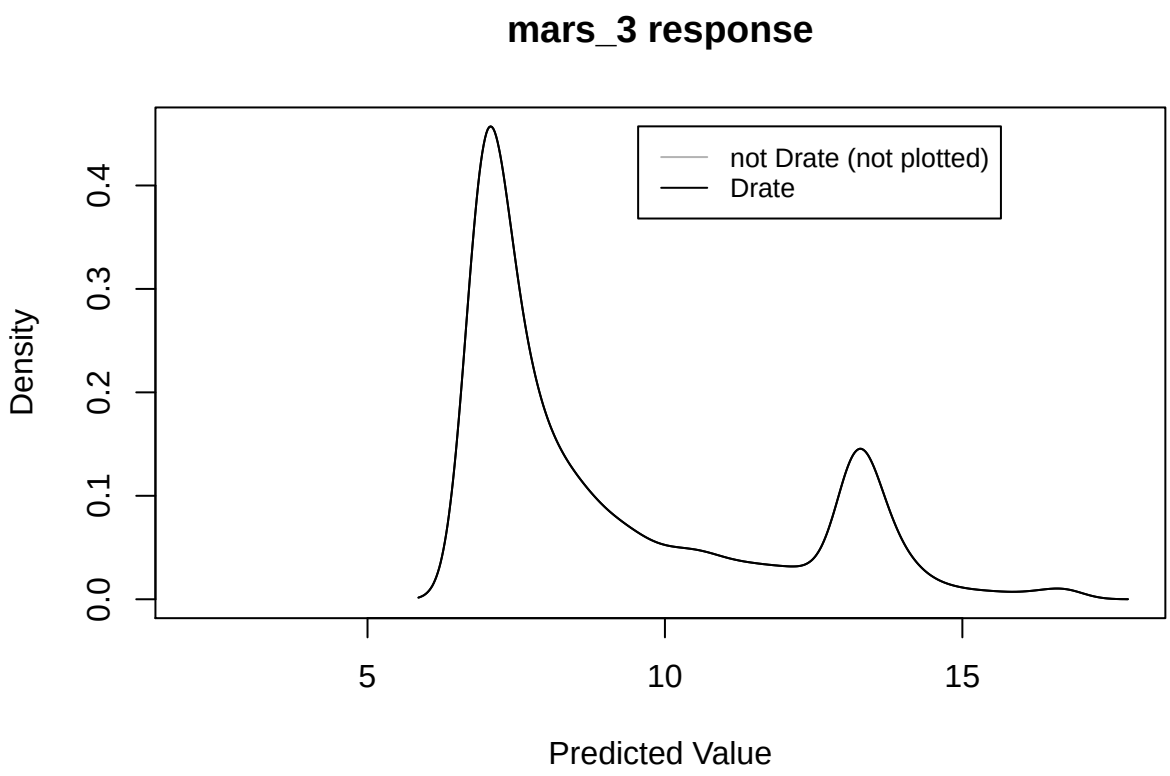
```
# MARS model on 'Num' feature
mars_3 <- earth(Drate ~ Num, data = default_rates, pmethod="backward", nprune=20, nfold=10);
summary(mars_3, digit=3);
```

```
## Call: earth(formula=Drate~Num, data=default_rates, pmethod="backward",
##             nprune=20, nfold=10)
##
##             coefficients
## (Intercept)       13.044
## h(140-Num)        -0.045
## h(Num-140)         0.003
## h(Num-1404)       -0.003
##
## Selected 4 of 4 terms, and 1 of 1 predictors
## Termination condition: RSq changed by less than 0.001 at 4 terms
## Importance: Num
## Number of terms at each degree of interaction: 1 3 (additive model)
```

```
## GCV 42.6  RSS 978269  GRSq 0.135  RSq 0.136  CVRSq 0.136
##
## Note: the cross-validation sd's below are standard deviations across folds
##
## Cross validation:   nterms 4.00 sd 0.00    nvars 1.00 sd 0.00
##
##      CVRSq   sd     MaxErr   sd
##      0.136 0.01       55.3 5.23
```

```
plotd(mars_3)
```

```
## Warning: standard deviation of 'not Drate' density is 0, density is
## degenerate?
```



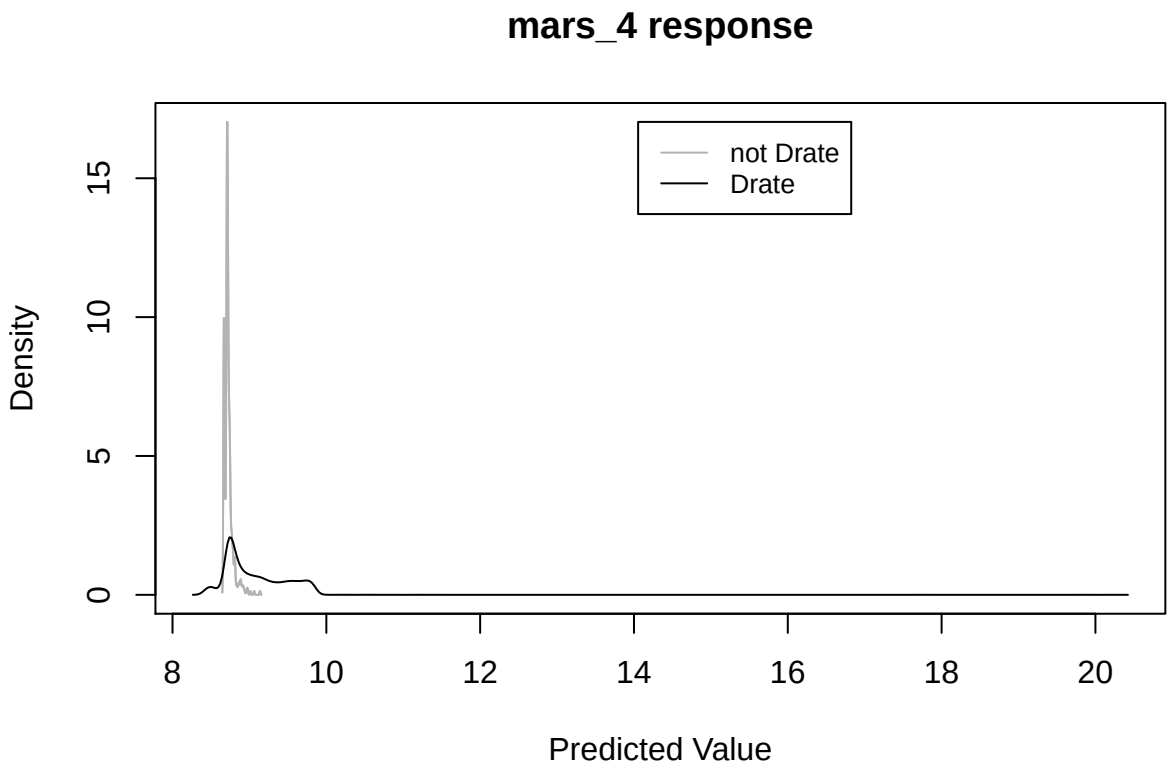
Model evaluation on 'Denom' feature

```
# MARS model on 'Denom' feature
mars_4 <- earth(Drate ~ Denom, data = default_rates, pmethod="backward", nprune=20, nfold=10);
summary(mars_4, digit=3);
```

```
## Call: earth(formula=Drate~Denom, data=default_rates, pmethod="backward",
##             nprune=20, nfold=10)
##
##               coefficients
## (Intercept)          9.858
## h(1016-Denom)       -0.001
```

```
## h(Denom-1016)       -0.001
## h(Denom-3866)        0.001
##
## Selected 4 of 4 terms, and 1 of 1 predictors
## Termination condition: RSq changed by less than 0.001 at 4 terms
## Importance: Denom
## Number of terms at each degree of interaction: 1 3 (additive model)
## GCV 49.1  RSS 1127974  GRSq 0.00293  RSq 0.00345  CVRSq 0.00175
##
## Note: the cross-validation sd's below are standard deviations across folds
##
## Cross validation:   nterms 3.60 sd 1.43    nvars 0.80 sd 0.42
##
##      CVRSq    sd     MaxErr   sd
##      0.002 0.002       53.8 6.76
```

```
plotd(mars_4)
```



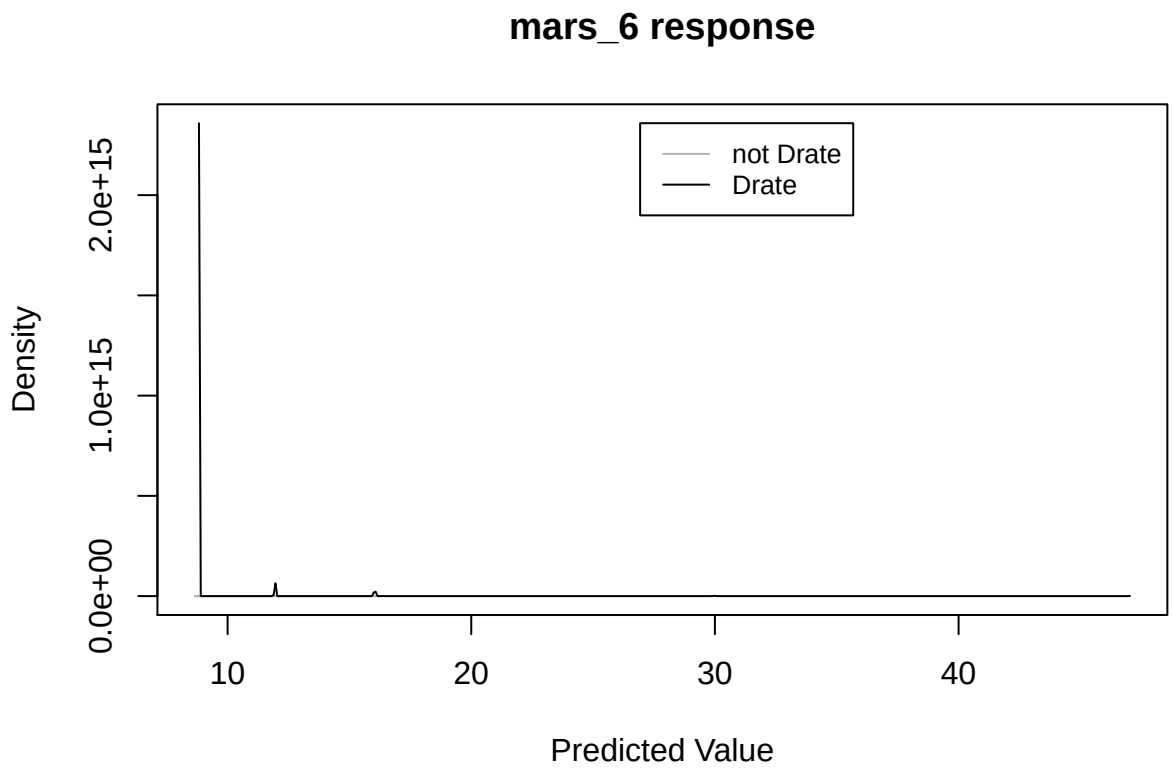
Model evaluation on 'EthnicCode' feature

```
# Linear model on 'EthnicCode' feature
mars_6 <- earth(Drate ~ EthnicCode, data = default_rates, pmethod="backward", nprune=20, nfold=10);
summary(mars_6, digit=3);
```

```
## Call: earth(formula=Drate~EthnicCode, data=default_rates,
```

```
##             pmethod="backward", nprune=20, nfold=10)
##
##             coefficients
## (Intercept)         30.0
## EthnicCode2        -13.9
## EthnicCode3        -18.0
## EthnicCode5        -21.1
## EthnicCodeD         17.1
##
## Selected 5 of 5 terms, and 4 of 4 predictors
## Termination condition: RSq changed by less than 0.001 at 5 terms
## Importance: EthnicCode5, EthnicCodeD, EthnicCode2, EthnicCode3
## Number of terms at each degree of interaction: 1 4 (additive model)
## GCV 47.4  RSS 1088300  GRSq 0.0378  RSq 0.0385  CVRSq 0.0378
##
## Note: the cross-validation sd's below are standard deviations across folds
##
## Cross validation:   nterms 5.00 sd 0.00    nvars 4.00 sd 0.00
##
##      CVRSq    sd     MaxErr   sd
##      0.038 0.014       53.7 7.14
```

```
plotd(mars_6)
```



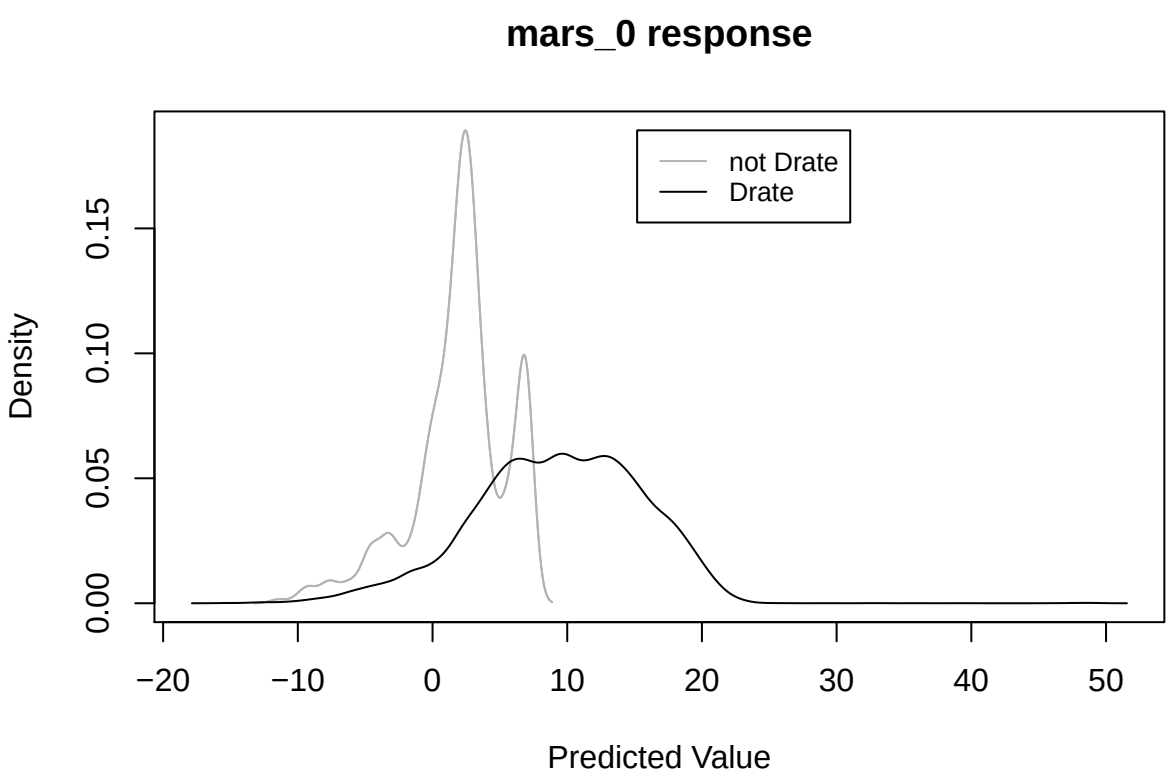
Final MARS Model

MARS model with multiple features

```
# MARS model on multiple features
mars_0 <- earth(Drate ~ ProgLength + SchoolType + Num + Denom + EthnicCode, data = default_rates, pmeth
summary(mars_0, digit=3);
```

```
## Call: earth(formula=Drate~ProgLength+SchoolType+Num+Denom+EthnicCode,
##             data=default_rates, pmethod="backward", nprune=20, nfold=10)
##
##               coefficients
## (Intercept)         25.196
## ProgLength3         -0.149
## ProgLength6         -0.722
## ProgLength8         -1.408
## EthnicCodeD         27.566
## h(Num-8)            -1.250
## h(Num-23)           -0.263
## h(Num-62)           -0.069
## h(141-Num)          -1.675
## h(Num-141)           1.612
## h(Num-607)          -0.025
## h(Denom-112)         0.054
## h(Denom-262)         0.022
## h(Denom-866)         0.008
## h(2446-Denom)        0.089
## h(Denom-2446)       -0.086
## h(Denom-7632)        0.001
##
## Selected 17 of 17 terms, and 6 of 21 predictors
## Termination condition: RSq changed by less than 0.001 at 17 terms
## Importance: Num, Denom, ProgLength8, EthnicCodeD, ProgLength6, ...
## Number of terms at each degree of interaction: 1 16 (additive model)
## GCV 10.2  RSS 234108  GRSq 0.793  RSq 0.793  CVRSq 0.793
##
## Note: the cross-validation sd's below are standard deviations across folds
##
## Cross validation:   nterms 17.50 sd 0.97    nvars 6.20 sd 0.63
##
##      CVRSq    sd     MaxErr   sd
##      0.793 0.011       44.3 8.49
```

```
plotd(mars_0)
```


mars_0 response
not Drate
Drate
Density
0.00
0.05
0.10
0.15
−20
−10
0
10
20
30
40
50
Predicted Value