Multivariate Adaptive Regression Spline (MARS)

Chapter 7-Part II Regression Models

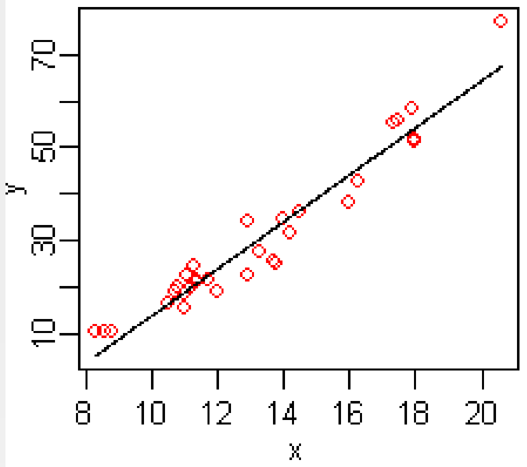
MARS

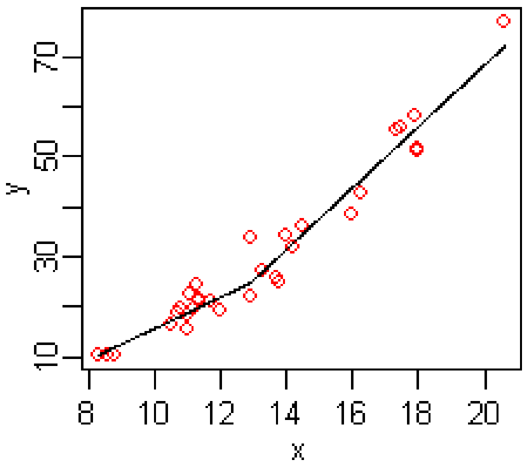
- MARS is a form of *stepwise linear regression*, which can be viewed as an extension of linear model that can model non-linearities.
- Introduced by Jerome Friedman in 1991.
- The term MARS is trademarked and licensed to Salford Systems.
- Suitable for higher dimensional inputs.
- MARS models are simpler as compared to other models like neural networks or random forest.

Terminology

- *Multivariate* - able to generate model based on several input variables (high dimensionality).
- *Adaptive* - generates flexible models in passes each time adjusting the model.
- *Regression* - estimation of relationship among independent and dependent variables.
- *Spline* - a piecewise defined polynomial function that is smooth (possesses higher order derivatives) where polynomial pieces connect.
- *Knot (tuning parameter)* - the point at which two polynomial pieces connect.

Linear regression vs MARS





Linear regression

MARS

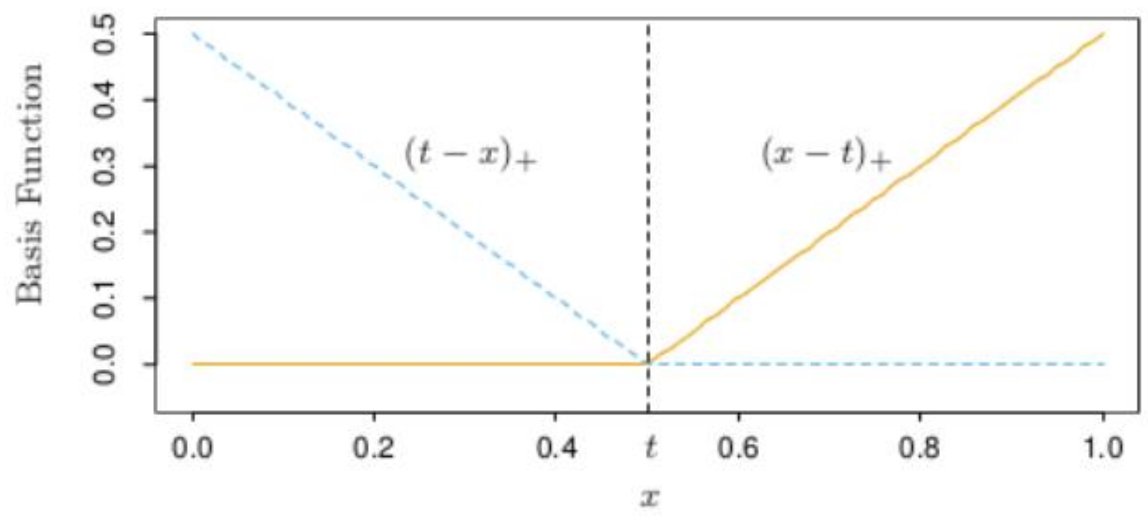
$y' = -37 + 5.1x$

$y' = 25 + 6.1 \max(0, x-13) - 3.1 \max(0, 13-x)$

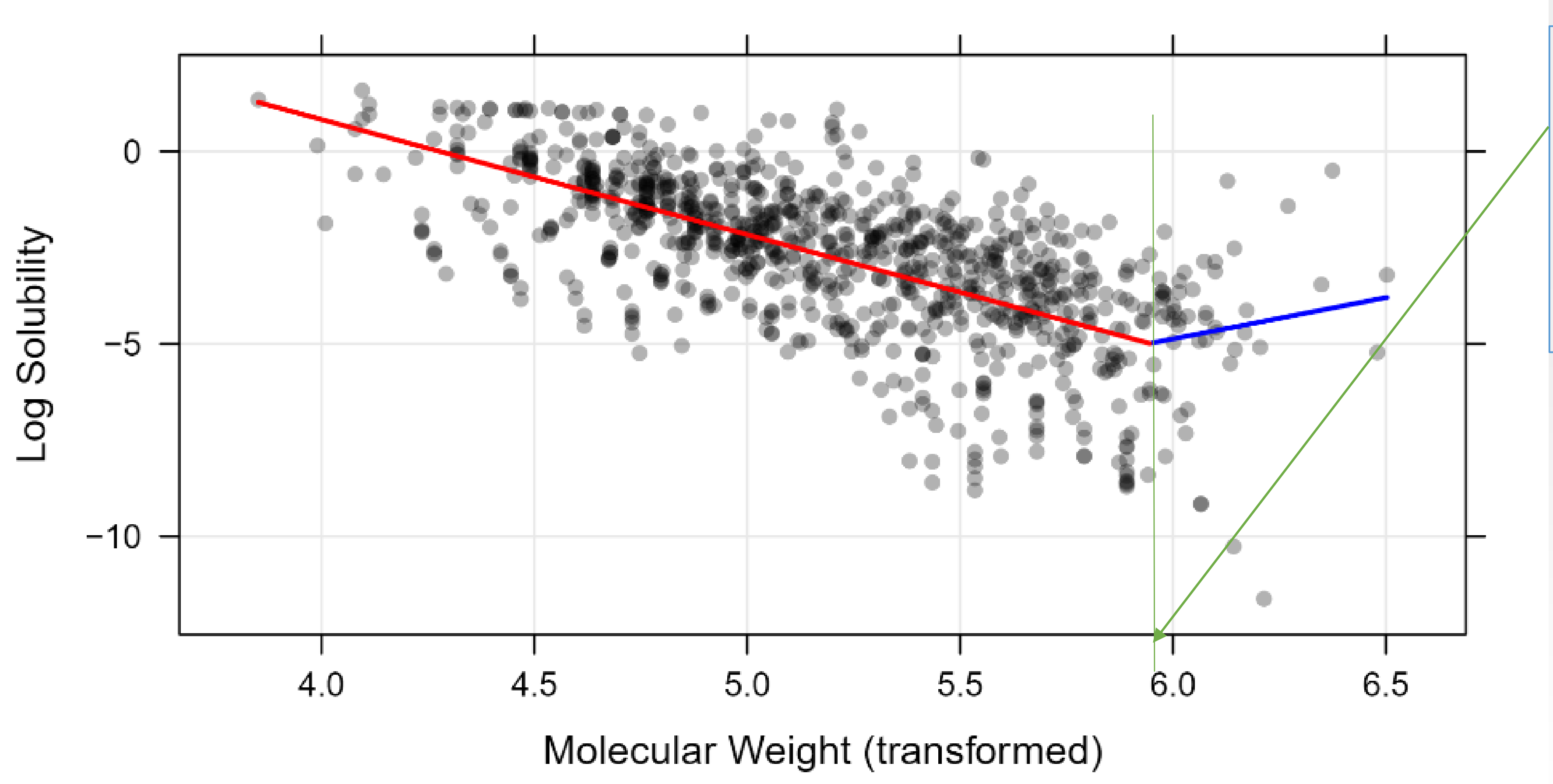
Basis functions

- MARS uses piecewise linear basis functions of the form $(x\text{-}t)_+$ and $(t\text{-}x)_+$. The `+' means positive part only. So

$$(x-t)_+ = \begin{cases} x-t, & \text{if } x>t, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad (t-x)_+ = \begin{cases} t-x, & \text{if } x<t, \\ 0, & \text{otherwise.} \end{cases}$$



MARS for Molecular Weight in the solubility data



- Question: how was the cut point (*knot*) determined?

MARS

- MARS model has the general form

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(x)$$

- $h_m(x)$ is a function from set of candidate functions or a product of two or more such functions.
- The coefficients β are estimated by minimizing the residual sum of squares (standard linear regression).
- These coefficients can be considered *weights* that represent the importance of the variable.

Remarks

- The model *automatically* conducts feature selection.
- The model could provide quantify the importance of each predictor to the model.
- The model provides clear interpretations of how each predictor relates to the outcome.
- The model requires very little pre-processing of the data; data transformations, near zero variance, and the filtering of predictors are not needed.
- Correlated predictors do not drastically affect model performance, but they can complicate model interpretation.

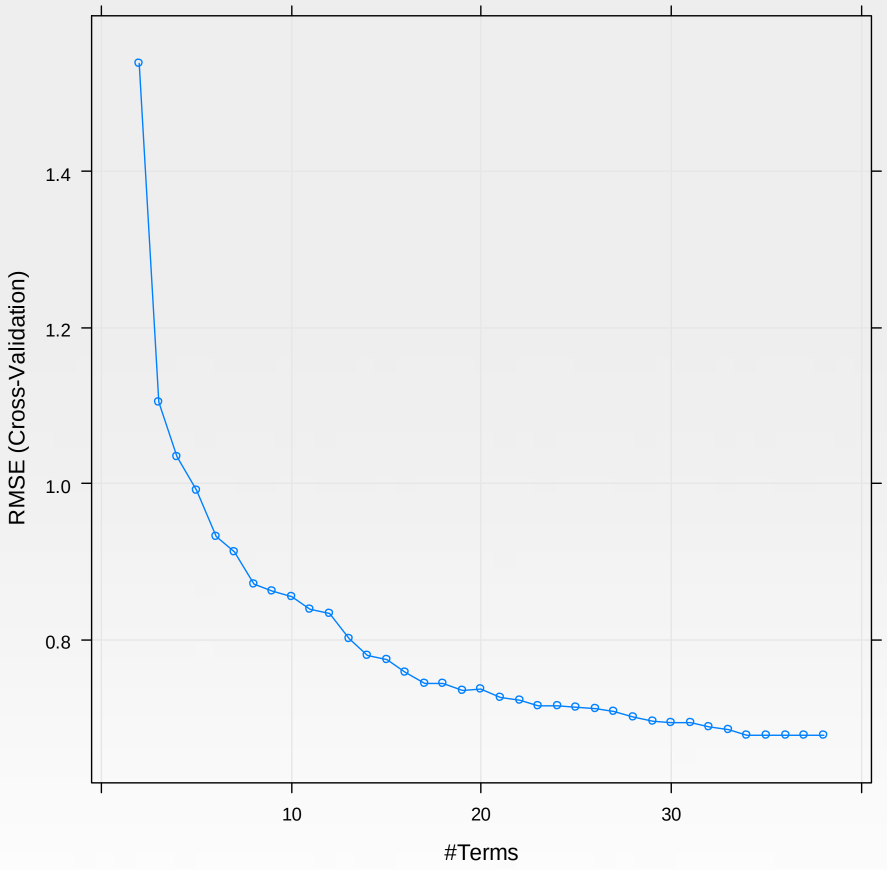
R codes for MARS

```
### Multivariate Adaptive Regression Splines
ptm <- proc.time() #takes 163 seconds to run in my computer
set.seed(100)
marsTune <- train(x = solTrainXtrans, y = solTrainY,
                  method = "earth",
                  tuneGrid = expand.grid(degree = 1, nprune = 2:38),
                  trControl = ctrl)
marsTune
proc.time() – ptm

plot(marsTune)

#Check the importance of each predictor
marsImp <- varImp(marsTune, scale = FALSE)
plot(marsImp, top = 25)

#save the predicted values into testResults
testResults$MARS <- predict(marsTune, solTestXtrans)
```

Tuning parameter

```
> marsTune
Multivariate Adaptive Regression Spline

951 samples
228 predictors

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 856, 855, 857, 856, 856, 855, ...
Resampling results across tuning parameters:

  nprune  RMSE       Rsquared   MAE
   2      1.5390057  0.4375015  1.1745616
   3      1.1044691  0.7125038  0.8408548
   4      1.0354182  0.7457278  0.7977694
   5      0.9920646  0.7659750  0.7521363
   6      0.9324996  0.7918923  0.7176847
   .
   .
   .
  35      0.6781610  0.8893081  0.5185573
  36      0.6772147  0.8895992  0.5181628
  37      0.6778717  0.8893533  0.5185987
  38      0.6774625  0.8893991  0.5176233

Tuning parameter 'degree' was held constant at a value of 1
RMSE was used to select the optimal model using the smallest value.
The final values used for the model were nprune = 36 and degree = 1.
```

Tuning parameter



Feature importance