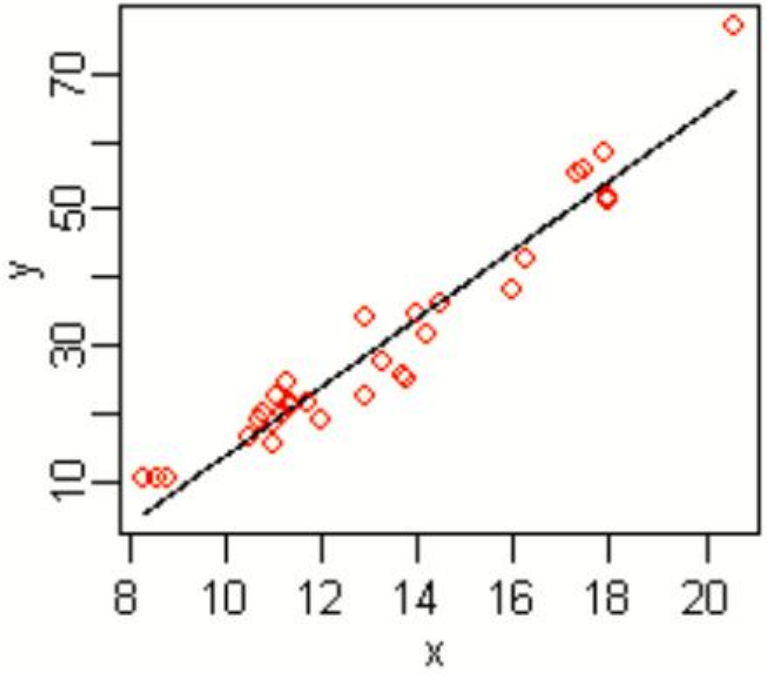
Advanced Regression Techniques

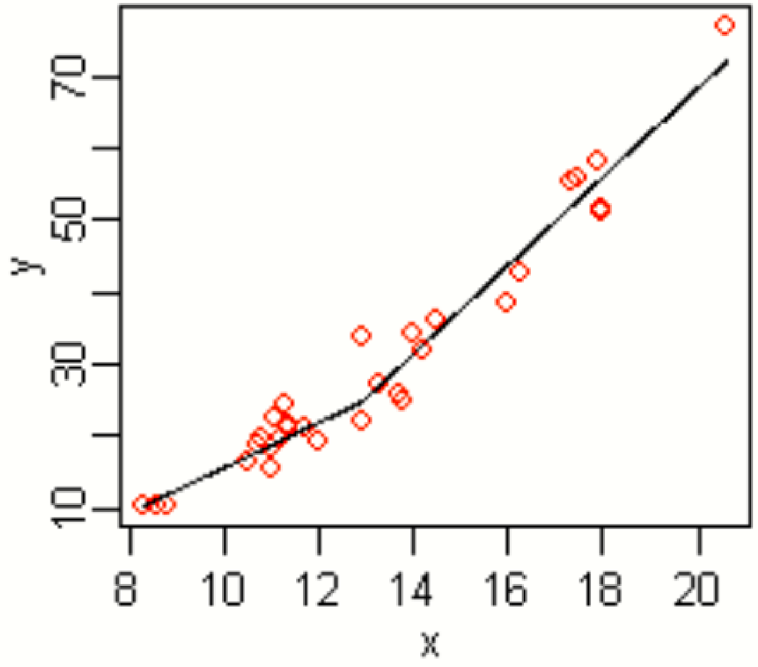
Multivariate Adaptive Regression Splines (MARS)

MARS

- Form of stepwise linear regression.
- Introduced by Jerome Friedman in 1991.
- The term MARS is trademarked
- Suitable for higher dimensional inputs.
- Extension of linear model that can model non-linearities.
- Modification of CART to improve regression performance
- MARS models are simpler as compared to other models like random forest or neural networks.

OLS vs MARS





Normal Regression

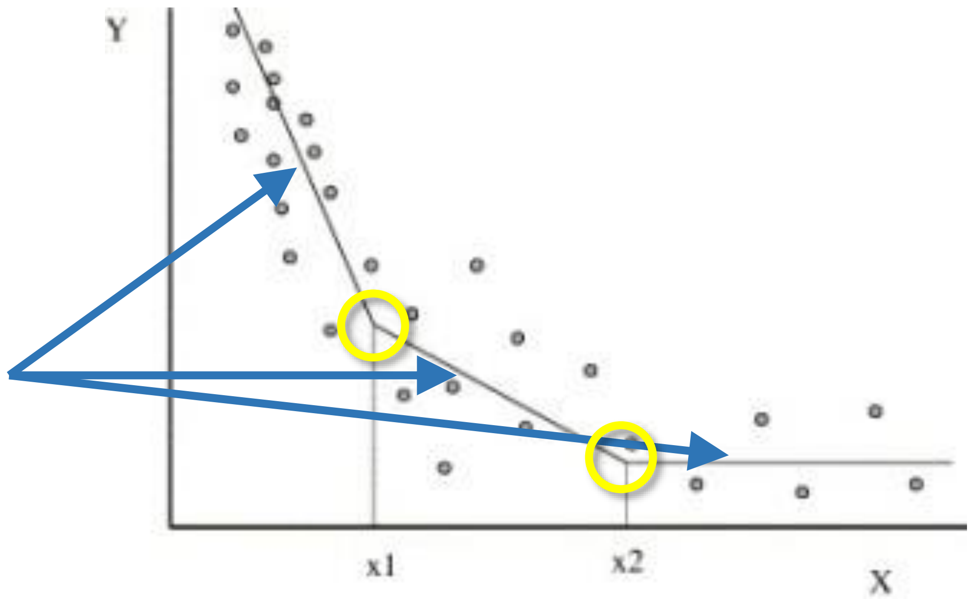
MARS

$y' = -37 + 5.1x$

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

Terminology

- Multivariate – multiple input variables
- Adaptive – generates flexible models in passes; each time adjusting the model
- Spline – Piecewise defined polynomial function that is smooth (possesses higher order derivatives) where polynomial pieces connect
- Knot - point at which two polynomial pieces connect



Basis functions

MARS uses piecewise linear basis functions of the form $(x-t)_+$ and $(t-x)_+$

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

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

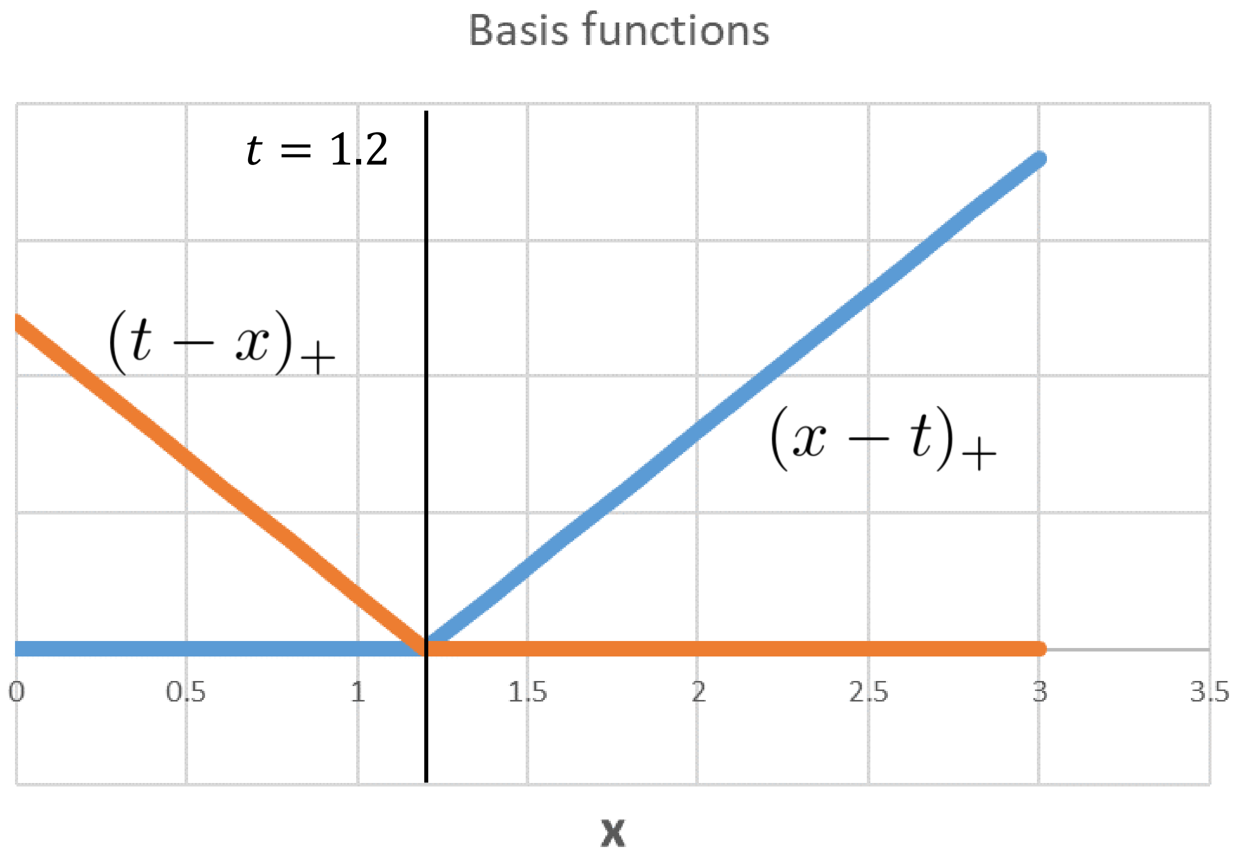
Basis functions

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

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

$$\text{Assume } x \in [0,3], \text{ assume } t = 1.2$$

x
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6
2.8
3

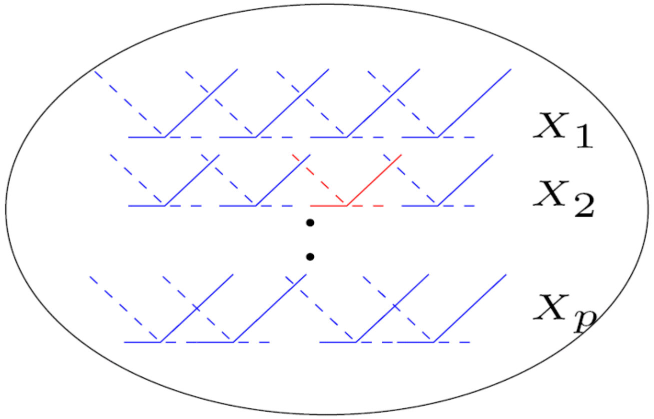


Basis functions

- MARS uses collections of functions comprised of reflected pairs for each variable X_j with knots at each observed value x_{ij} of that variable.
- So, for p predictors with n observations, the *candidate* set is:

$$\mathcal{C} = \{(X_j - t)_+, (t - X_j)_+\}$$

$$\forall \begin{array}{l} t \in \{x_{1j}, x_{2j}, \dots, x_{nj}\} \\ j = 1, 2, \dots, p \end{array}$$



- If all of the input values are distinct, then C contains $2np$ functions!

Candidate set example

Predictors		Target
X1	**X2**	**Y**
0.78	92.13	10.65
1.02	39.56	10.19
4.28	25.69	25.08
5.22	35.77	34.39
5.45	15.34	14.79

$$
\begin{aligned}
\mathcal{C} = \{&(X_1 - 0.78)_+, (0.78 - X_1)_+,\\
&(X_1 - 1.02)_+, (1.02 - X_1)_+,\\
&(X_1 - 4.28)_+, (4.28 - X_1)_+,\\
&(X_1 - 5.22)_+, (5.22 - X_1)_+,\\
&(X_1 - 5.45)_+, (5.45 - X_1)_+,\\
&(X_2 - 92.13)_+, (92.13 - X_2)_+,\\
&(X_2 - 39.56)_+, (39.56 - X_2)_+,\\
&(X_2 - 25.69)_+, (25.69 - X_2)_+,\\
&(X_2 - 35.77)_+, (35.77 - X_2)_+,\\
&(X_2 - 15.34)_+, (15.34 - X_2)_+\}
\end{aligned}
$$

MARS model equation

- General form:

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

- $h_m(X)$ is a function from set $\mathcal{C}$ of candidate functions; or, a product of two or more such functions.
- Betas are the coefficients estimated by minimizing the residual sum of squares (OLS).

MARS model building process

- Calculate set of candidate functions $\mathcal{C}$ by generating reflected pairs of basis functions with knots set at observed values.
- Specify constraints: i.e., the number of terms in the model and maximum allowable degree of interaction.
- Do forward pass - try out new function products and see which product decreases training error.
- Do backward pass – to fix overfit.
- Do generalized cross validation to estimate the optimal number of terms in the model.

MARS forward pass

- At each step, MARS adds the basis function which reduces the residual error the most
  - this is almost identical to forward stepwise regression
- Always adds the basis function in 'pairs', both sides of knot
- Calculate value for knot and function that fit the data
- The addition of model terms continues until the *max number* of terms (pre-specified) in the model is reached

MARS backward pass

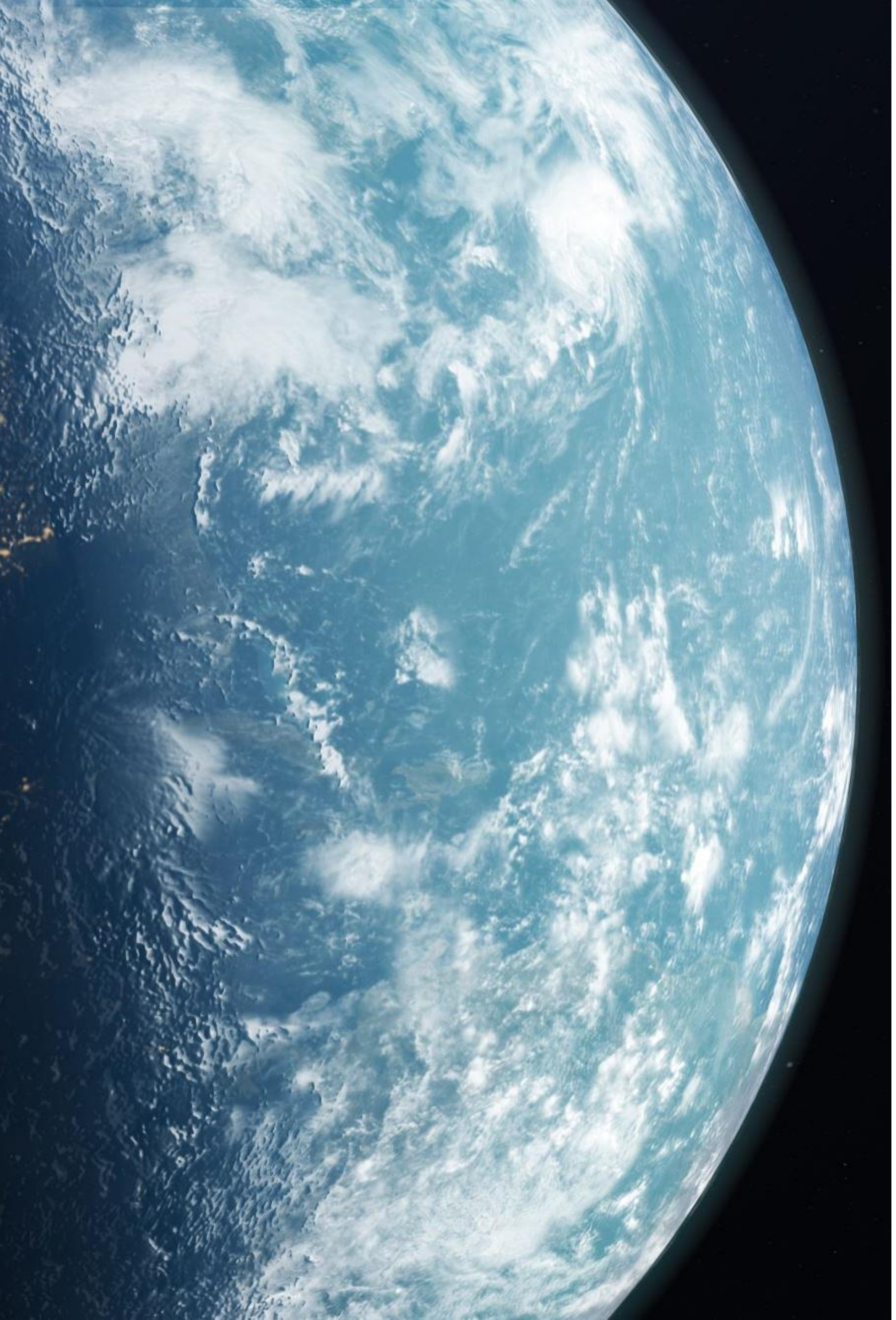
- Remove one term at a time from the model, i.e., the term which increases the residual error the least
  - Essentially, backward stepwise regression
- Continue removing terms until cross validation is satisfied
  - Use the Generalized Cross Validation (GCV) function for this purpose

Generalized Cross Validation (GCV)

- Metric used in the backwards stepwise regression step – similar in spirt to adjusted R^2 or AIC in that it penalizes the value based on the model complexity

$$\begin{aligned} GCV(\lambda) &= \sum_{i=1}^{N} \frac{(y_i - \hat{f}_\lambda(x_i))^2}{(1 - M(\lambda)/N)^2} \\ M(\lambda) &= r + c \cdot K \end{aligned}$$

- M() measures effective # of parameters:
  - r: # of linearly independent basis functions
  - K: # of knots selected
  - c = 3



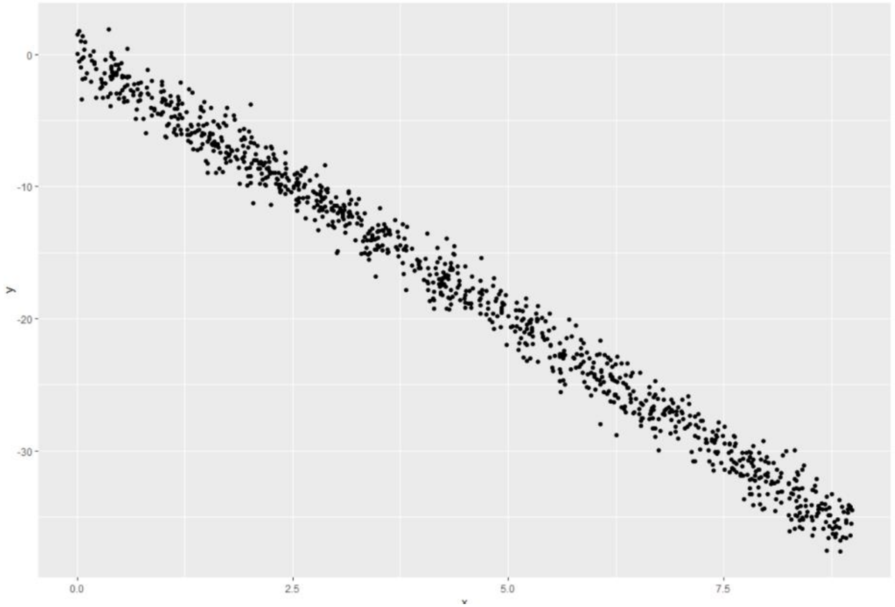
MARS example in R

- MARS is available in the `earth` library
- See the “mars Example.r” file in the course website

MARS example

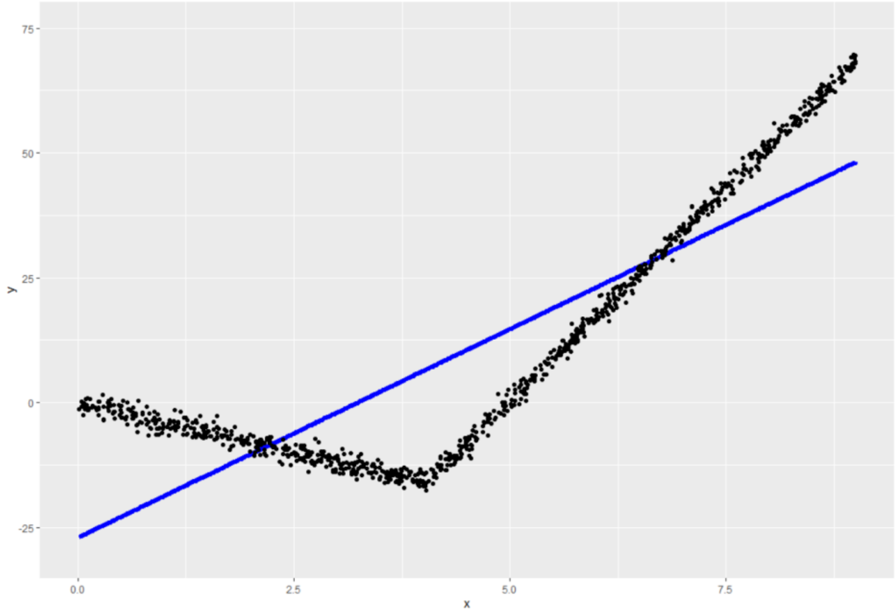
- Type of relationship expected for OLS

```
Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept)  0.10057    0.07550    1.332    0.183
x           -4.01300    0.01469 -273.090   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.21 on 998 degrees of freedom
Multiple R-squared:  0.9868,    Adjusted R-squared:  0.9868
F-statistic: 7.458e+04 on 1 and 998 DF,  p-value: < 2.2e-16
```



- But what if you have this?

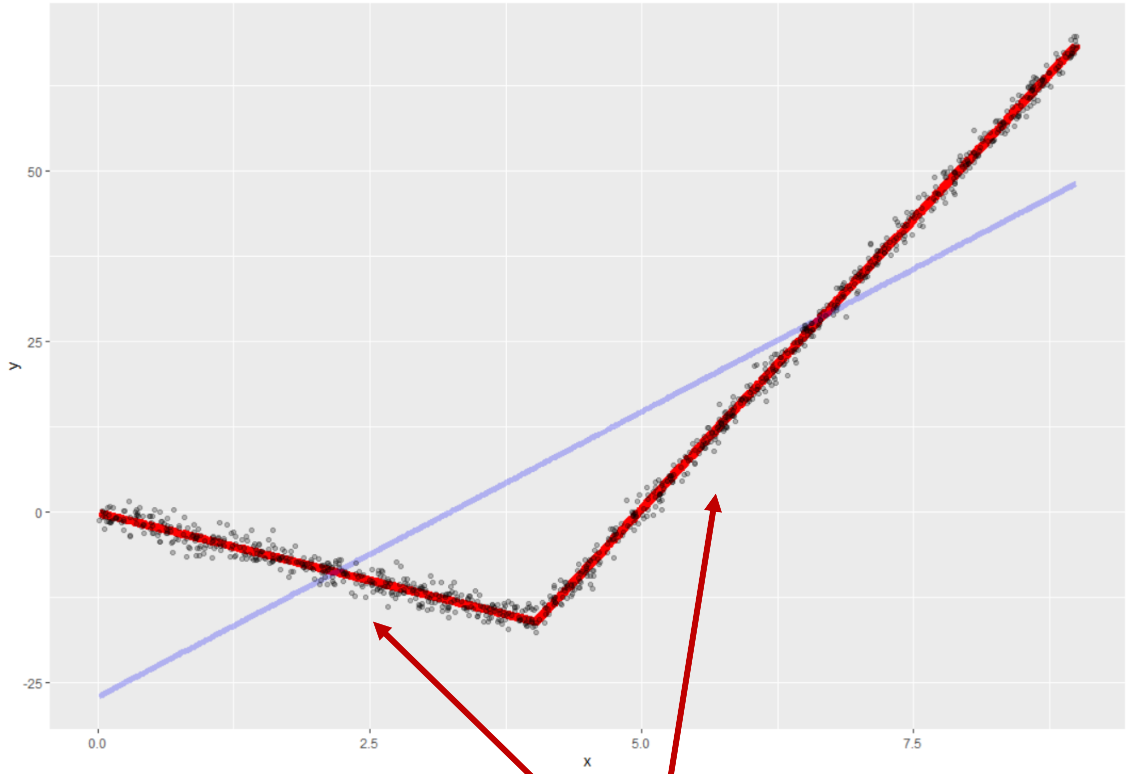
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -27.0844     0.8684  -31.19   <2e-16 ***
x             8.3612     0.1649   50.71   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.44 on 998 degrees of freedom
Multiple R-squared:  0.7204,    Adjusted R-squared:  0.7201
F-statistic:  2572 on 1 and 998 DF,  p-value: < 2.2e-16
```



So, let’s try MARS

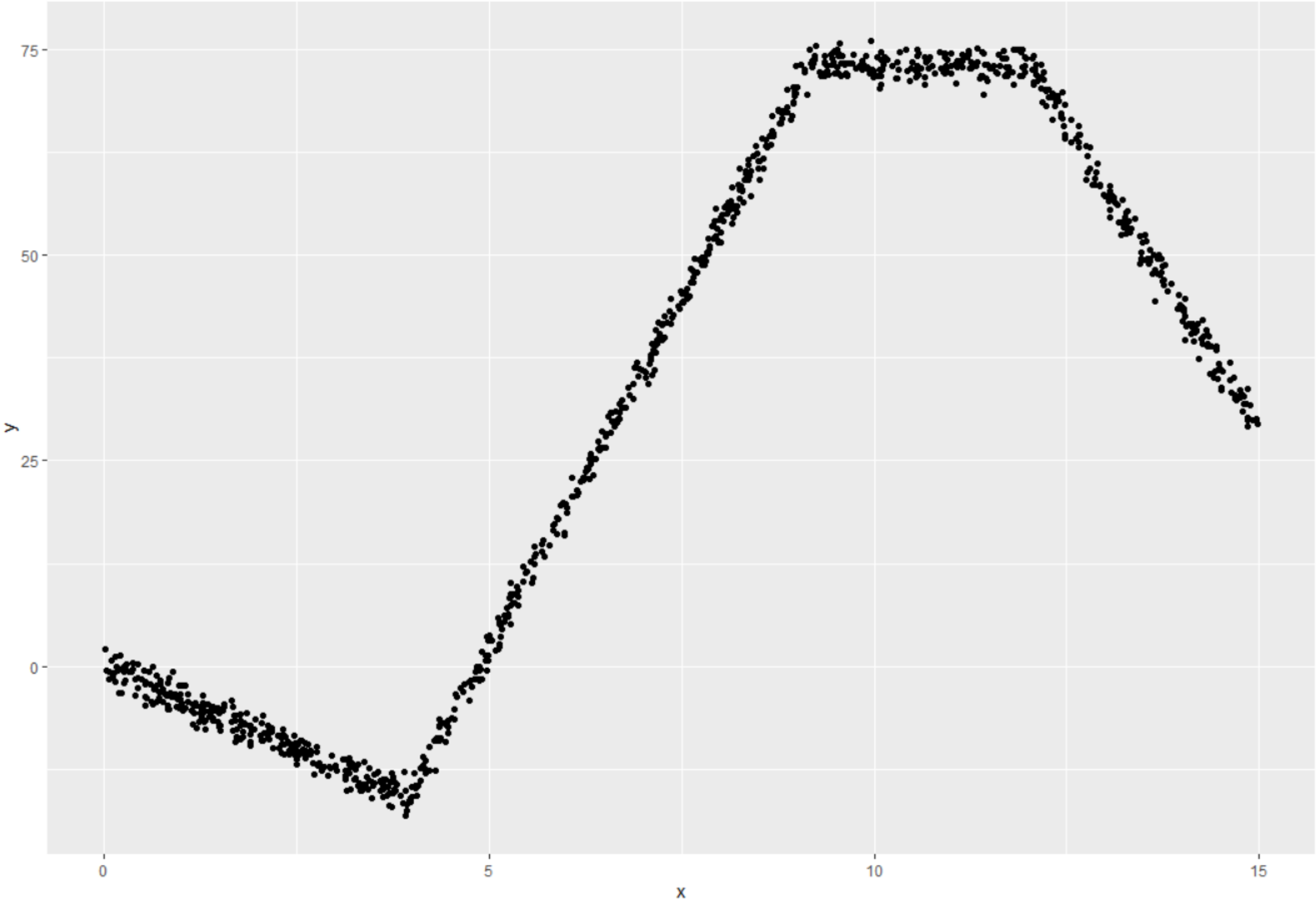
```
marsFit <- earth(y~x,data=df)
summary(marsFit, style="pmax")
```

```
Call: earth(formula=y~x, data=df)

y =
  -16.07977
  +  3.964837 * pmax(0, 4.028753 -        x)
  +  16.99043 * pmax(0,        x - 4.028753)

Selected 3 of 3 terms, and 1 of 1 predictors
Termination condition: RSq changed by less than 0.001 at 3 terms
Importance: x
Number of terms at each degree of interaction: 1 2 (additive model)
GCV 1.503868    RSS 1488.867    GRSq 0.9976718    RSq 0.9976904
```



Two slopes!
- for x < 4, slope ≈ 4
- for x > 4, slope ≈ 17

Now, for crazy time…



```
Call: earth(formula=y~x, data=df)

y =
  -60.65957
  +  18.11114 * pmax(0,        x - 3.383024)
  +  4.922696 * pmax(0,        x - 5.681345)
  -  18.41192 * pmax(0,        x - 9.207713)
  +  5.447523 * pmax(0, 11.42461 -        x)
  -   2.53239 * pmax(0,        x - 11.42461)
  -  16.96373 * pmax(0,        x - 11.97612)

Selected 7 of 7 terms, and 1 of 1 predictors
Termination condition: RSq changed by less than 0.001 at 7 terms
Importance: x
Number of terms at each degree of interaction: 1 6 (additive model)
GCV 3.820366    RSS 3721.682    GRSq 0.9965005    RSq 0.996584
```

but if greater than 3.38, then $slope \leftarrow -5.4 + 18.1 = 12.7$

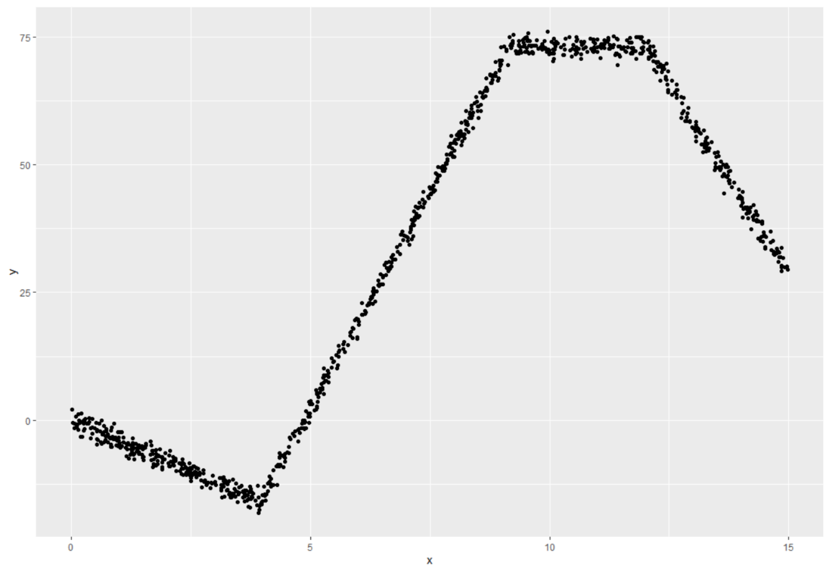
and if greater than 5.68, then $slope \leftarrow 12.7 + 4.9 = 17.6$

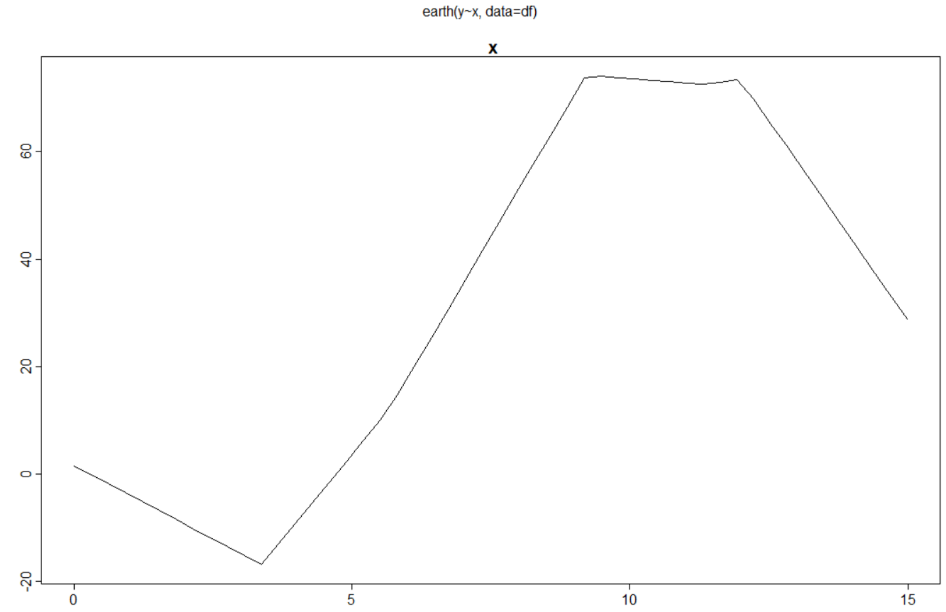
and if also greater than 9.2, then $slope \leftarrow 17.6 - 18.4 = -0.8$

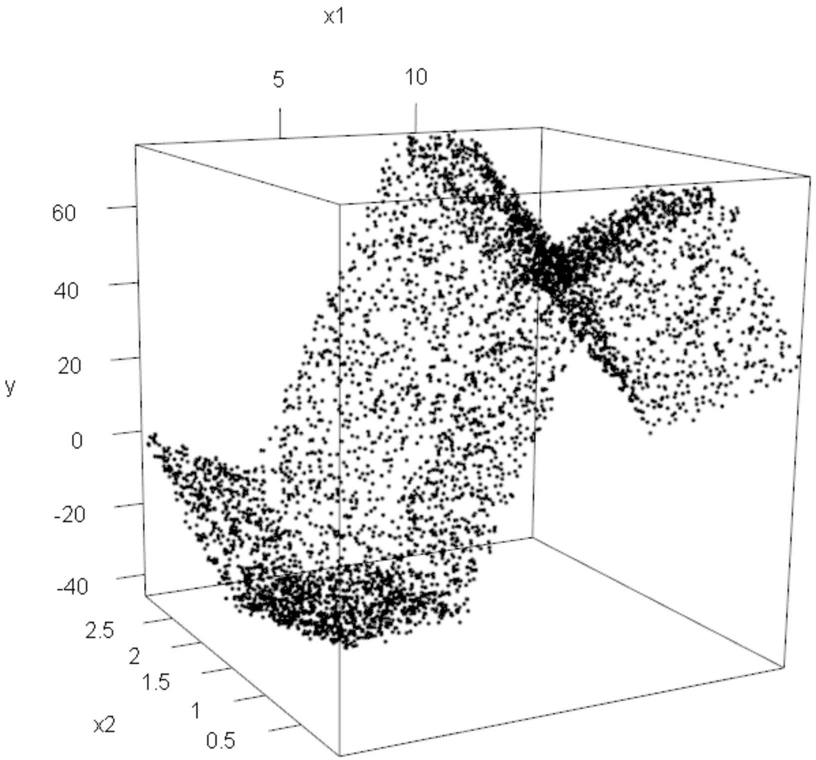
if x is less than 11.4, then $slope \leftarrow -5.4$

and if also greater than 11.4, then $slope \leftarrow -0.8 - 2.5 + 5.4 = 2.1$

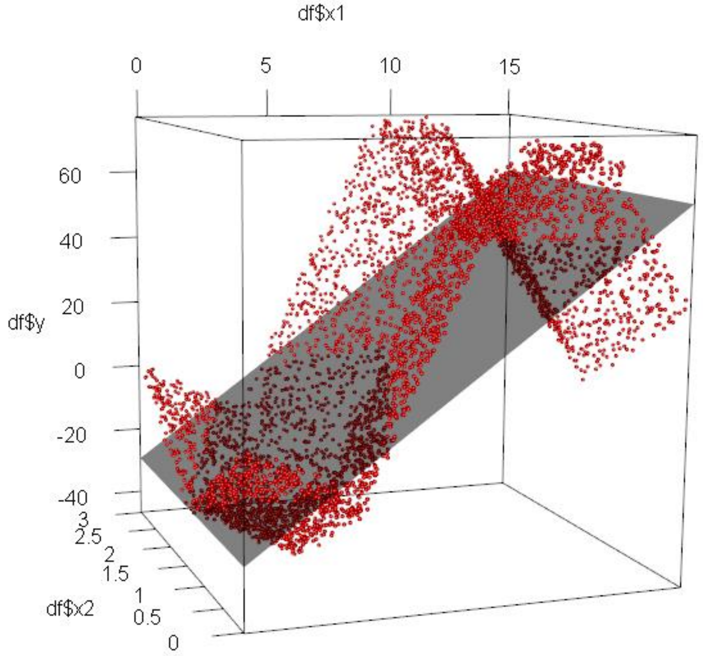
and if also greater than 11.98, then $slope \leftarrow 2.1 - 17.0 = -14.9$



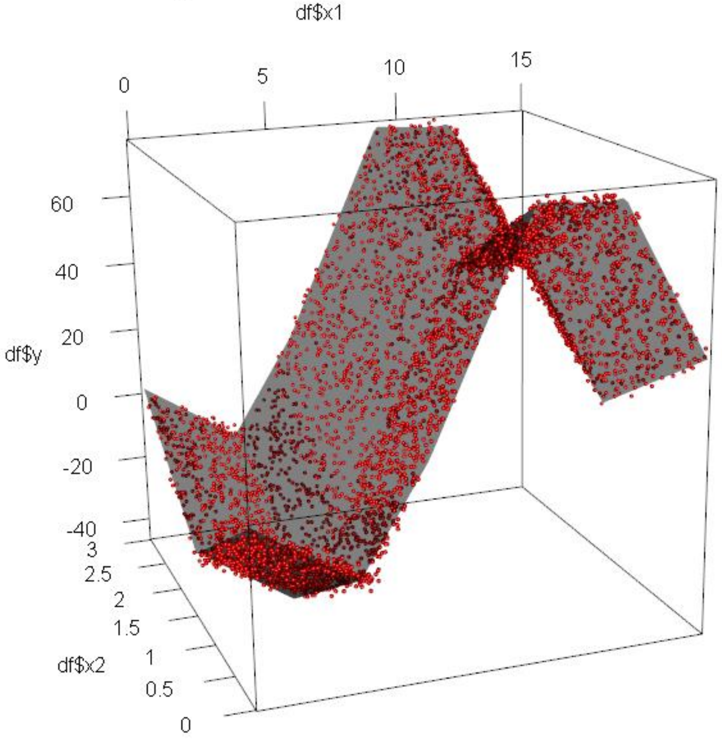


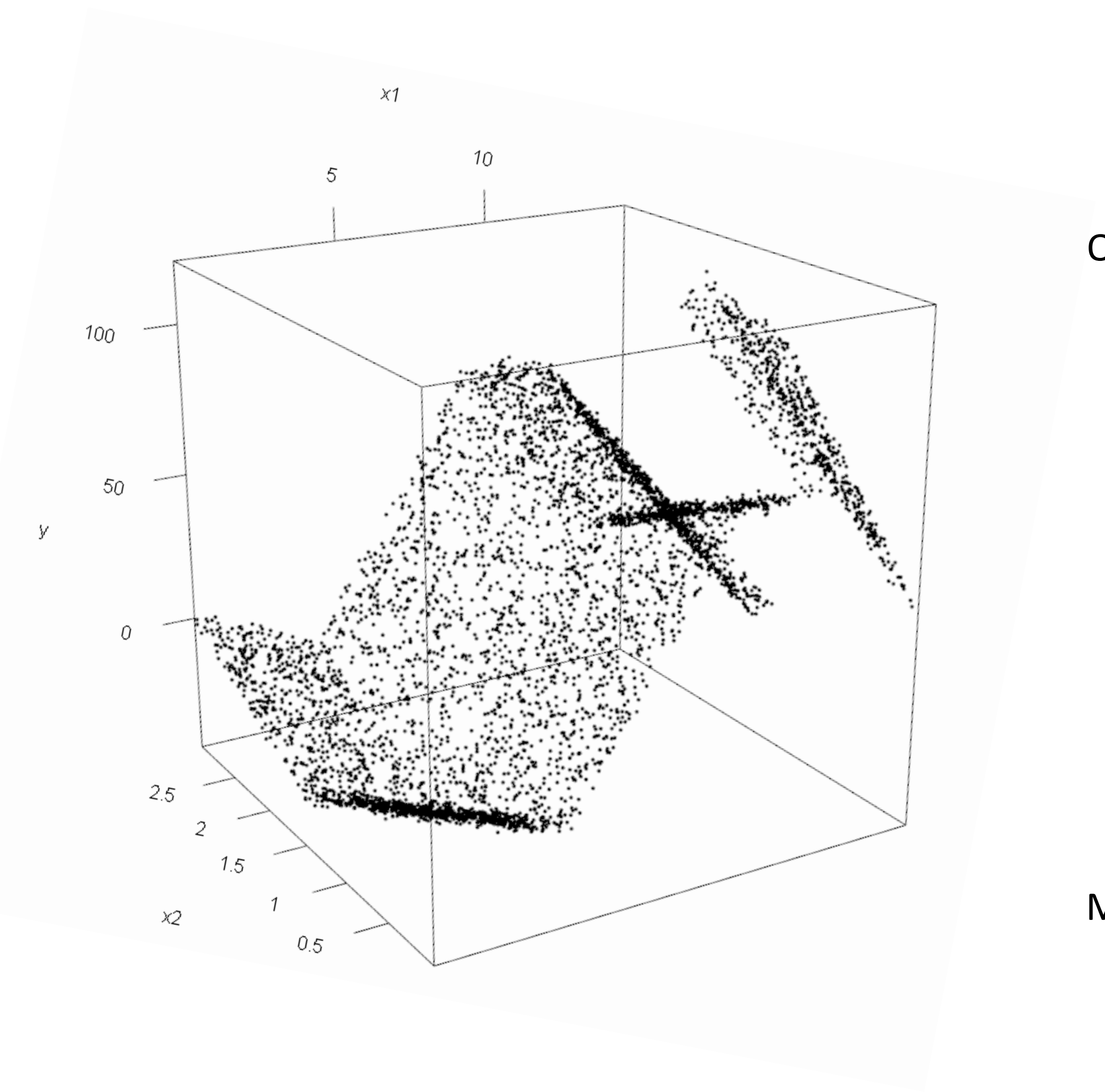


OLS fit →

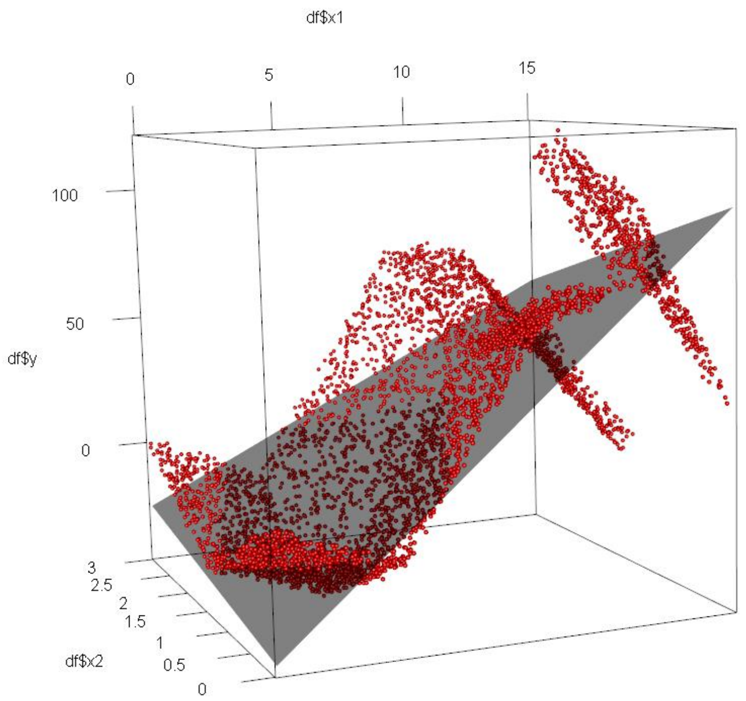


MARS fit →





OLS fit →



MARS fit →