Introduction to Multivariate Adaptive Regression Splines (MARS)

Randy Law
Analytics & Insights Matter
Highlands Ranch, CO

Philip Turk, PhD[1]

Western Data Analytics, LLC; Denver, CO

08/23/2019

[1] <pturk@westerndataanalytics.com>

Introduction

- Several procedures for fitting models are inherently *linear*. For example:
  - Simple and multiple (linear) regression
  - Logistic regression
- Nowadays, there are a variety of algorithms that are inherently non-linear.
- When using these models, the exact form of the non-linearity does not need to be known explicitly or specified prior to model fitting. Rather, these algorithms will search for, and discover, non-linearity in the data in a way that maximizes 'predictive accuracy'.
- This introduction discusses multivariate adaptive regression splines (MARS), an algorithm that essentially creates a piecewise linear model which provides an intuitive first step into non-linearity after grasping the concept of linear regression and other types of linear models.

Preliminaries

We will install and load the following packages:

```
> library(tidyverse)   # plotting
> library(rsample)     # data splitting
> library(earth)       # fit MARS models
> library(AmesHousing) # toy data set
```

Preliminaries

- To illustrate MARS, we will use famous Ames Housing data set.

```
> ?ames_raw
> ames <- make_ames() ## Create a clean version of the data
```

- Study the help page before you do anything else.
- Now go the Environment window and examine the data set.

Preliminaries

- As good machine learning modelers do, we create a *training* set from the data for modeling (70%) and a *test* set for assessing prediction (30%).
- We use `set.seed()` so that we can reproduce our results.

```
> set.seed(123) ## Pick your favorite
> ames_split <- initial_split(ames, prop = 0.7,
+                             strata = "Sale_Price")
> ames_train <- training(ames_split)
> ames_test  <- testing(ames_split)
```

Basic Background

- Linear models have advantages such as their ease and speed of computation and also the intuitive nature of interpreting their coefficients. However, linear models make a strong assumption about linearity (with respect to the parameters), and this assumption is often a poor one, which can affect predictive accuracy.
- We can extend linear models to capture non-linear relationships in a couple pedestrian ways.

Basic Background

- *Polynomial regression* is a form of regression in which the relationship between the independent variable x and the dependent variable y is modeled as an n^{th} degree polynomial of x.
- For example, the equation below represents a polynomial regression function where y is modeled as a function of x to the 2nd degree:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

- Generally speaking, it is unusual to use a degree greater than 3. The reason is because the larger the degree, the easier the function fit becomes overly flexible ("overfitting") and oddly shaped, especially near the boundaries of the domain of x.

Basic Background

- An alternative to polynomial regression is so-called *step function regression*.
- Whereas polynomial functions impose a global non-linear relationship, step functions break the domain of x into "local" bins, and fit a different constant for each bin. This amounts to converting a continuous variable into an ordered categorical variable such that our linear regression function is converted to the following equation:

$$y = \beta_0 + \beta_1 C_1(x) + \beta_2 C_2(x) + \cdots + \beta_d C_d(x) + \varepsilon$$

where so-called *indicator functions* $C_1(x)$ represents x values such that $c_1 < x \leq c_2$, $C_2(x)$ represents x values such that $c_2 < x \leq c_3$, and $C_d(x)$ represents x values such that $c_d < x$.

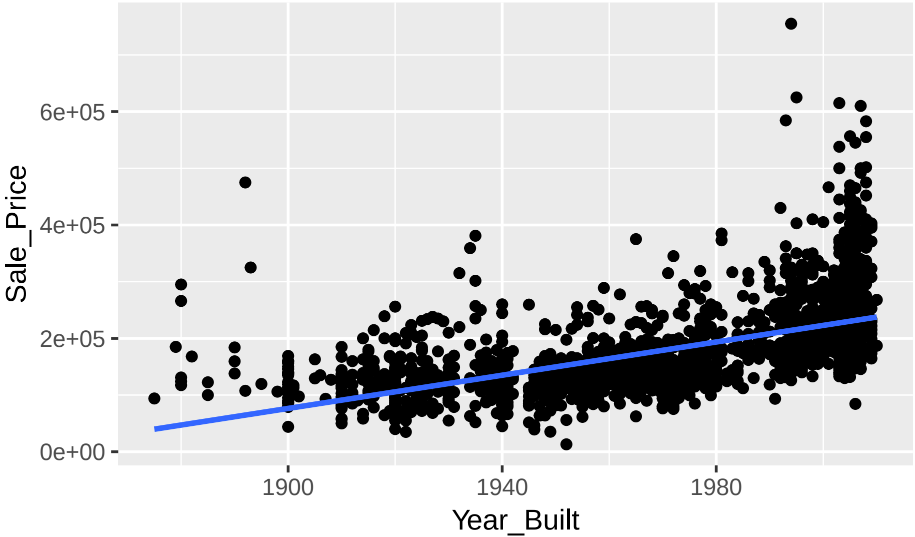
Basic Background

- Although useful, polynomial regression and step function regression require the user to explicitly identify and incorporate which variables should have what specific degree or at what points of a variable x should cut points be made for the step functions.
- Considering many data sets today can easily contain many, many variables, this would require an enormous, if not impossible, commitment to make these determinations.

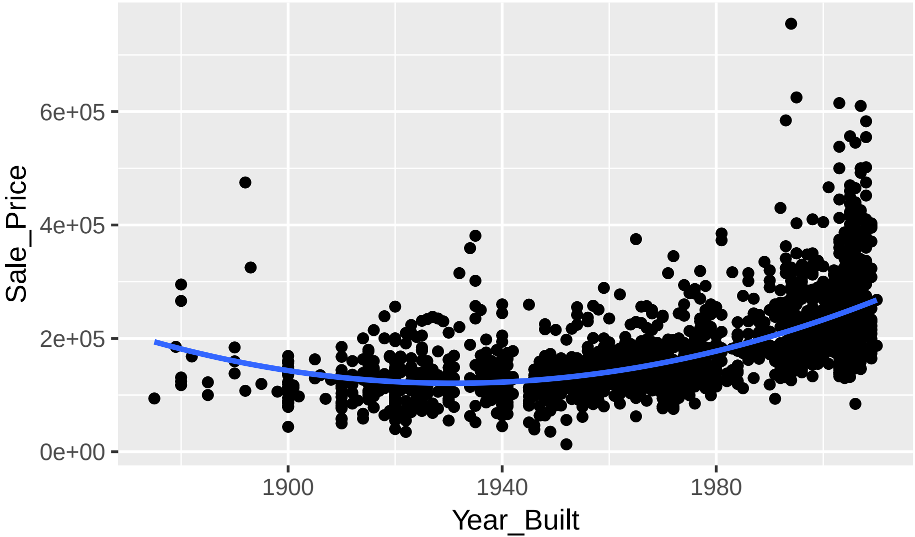
Exploratory Data Analysis: Graphs

- We now look at three graphs demonstrating the three previous models we have discussed. The blue line represents predicted `Sale_Price` values as a function of `Year_Built` for alternative approaches to modeling explicit non-linear regression patterns:
  - A. Traditional linear regression approach does not capture any non-linearity unless the predictor or response is transformed (i.e., log transformation),
  - B. Degree-2 polynomial,
  - C. Step function fitting cutting Year_Built into three categorical levels.

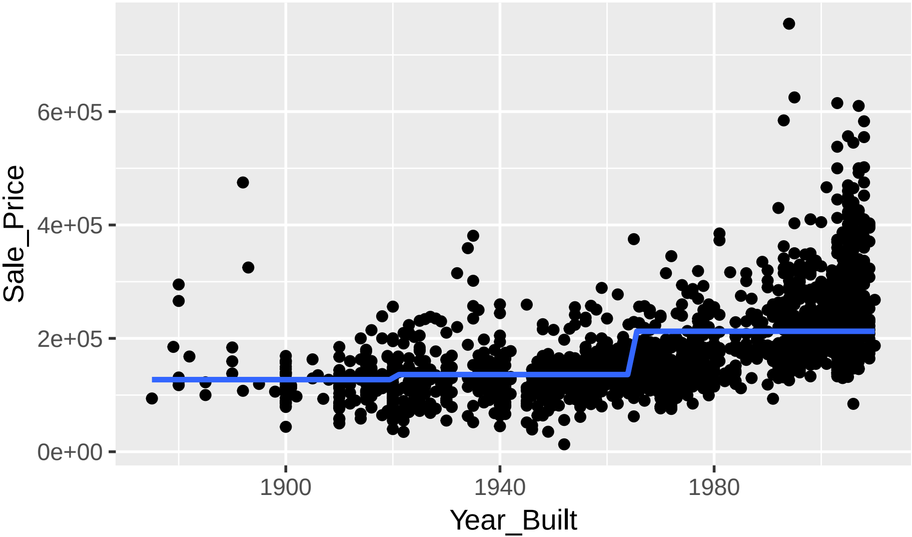
Exploratory Data Analysis: Graphs



Exploratory Data Analysis: Graphs



Exploratory Data Analysis: Graphs

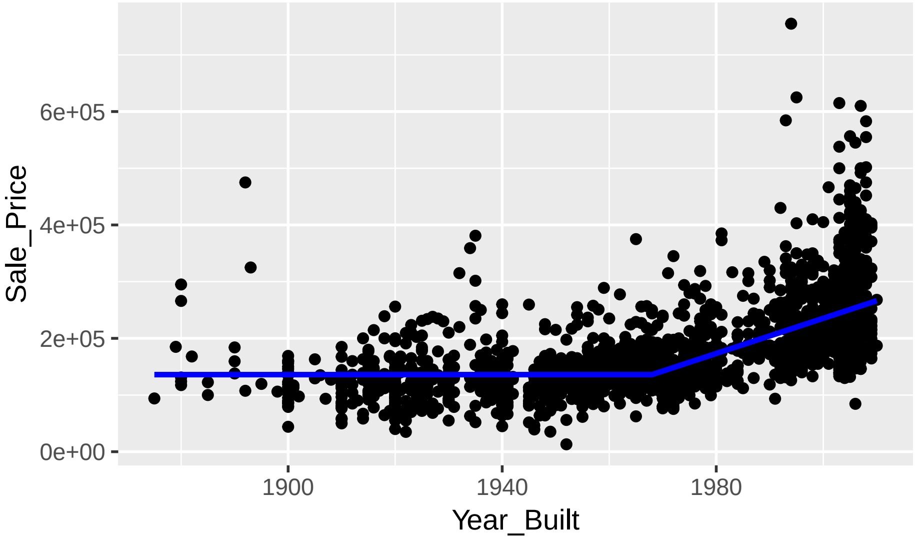


Multivariate Adaptive Regression Splines

- MARS provide a convenient approach to capture the non-linear flavor of polynomial regression by determining cutpoints (*knots*) similar to step functions, but using an algorithm.
- The procedure assesses each data point for each predictor as a potential knot and creates a linear regression model.
- For example, consider our simple model of `Sale_Price ~ Year_Built`. The MARS procedure will first look for the single point (knot) across the range of `Year_Built` values where two different linear relationships between `Sale_Price` and `Year_Built` achieve the smallest error. This creates a so-called 'hinge' effect at the knot:

$$\text{Sale\_Price} = \begin{cases} 135957.54 & \text{Year\_Built} \leq 1968 \\ 135957.54 + 3102.98(\text{Year\_Built} - 1968) & \text{Year\_Built} > 1968 \end{cases}$$

Multivariate Adaptive Regression Splines: Graph

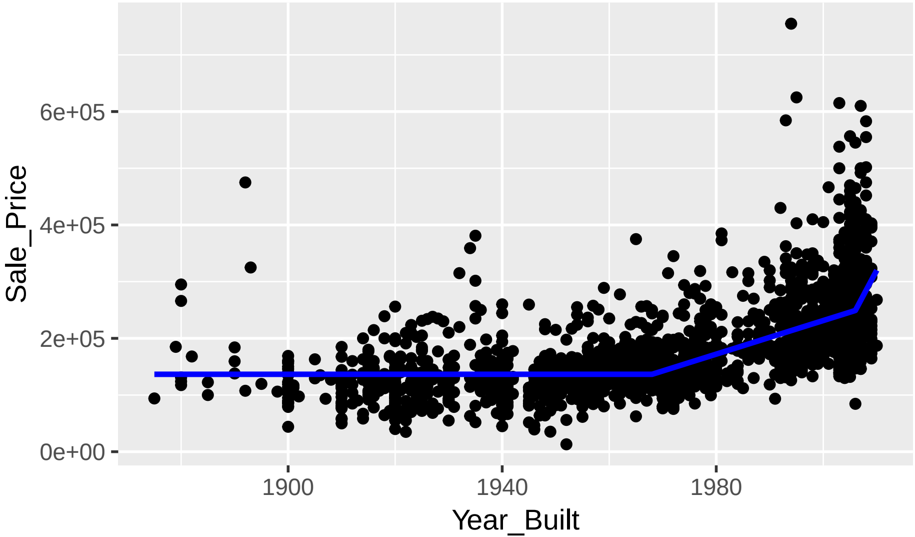


Multivariate Adaptive Regression Splines

- Once the first knot has been found, the search continues for a second knot which is found at 2006. This results in the following model:

$$\text{Sale\_Price} = \begin{cases} 136568.02 & \text{Year\_Built} \leq 1968 \\ 136568.02 + 2963.06(\text{Year\_Built} - 1968) & 1968 < \text{Year\_Built} \leq 2006 \\ 136568.02 + 2963.06(\text{Year\_Built} - 1968) & \text{Year\_Built} > 2006 \\ +14793.97(\text{Year\_Built} - 2006) & \end{cases}$$

Multivariate Adaptive Regression Splines: Graph



Multivariate Adaptive Regression Splines

- This procedure can continue until many knots are found, producing a highly non-linear pattern. Although including many knots may allow us to find/fit a really good relationship with our training set, it may not generalize very well to our test set.
- Consequently, once a sequence of knots have been created, we can sequentially remove knots that do not contribute significantly to predictive accuracy. This process is known as "pruning" and we can use *cross-validation* to find the optimal number of knots.

Fitting a Basic MARS Model

- We can fit a MARS model with the `earth` package. By default, `earth` will assess all potential knots across all supplied predictors and factors and then will prune to the optimal number of knots based on an expected change in a form of $R^2 < 0.001$.
- This calculation is performed by the generalized cross-validation procedure (GCV statistic).

Fitting a Basic MARS Model

- The following fits a basic MARS model to our `ames` training data and performs a search for required knots across all features. The results show us the final model's GCV statistic, generalized R^2 (GRSq), and more.

```
> mars01 <- earth(
+   Sale_Price ~ .,
+   data = ames_train
+ )
>
> mars01
```

```
Selected 32 of 36 terms, and 25 of 307 predictors
Termination condition: RSq changed by less than 0.001 at 36 terms
Importance: Gr_Liv_Area, Year_Built, Overall_QualExcellent, ...
Number of terms at each degree of interaction: 1 31 (additive model)
GCV 544054108    RSS 1.049444e+12    GRSq 0.9142289    RSq 0.9193337
```

Fitting a Basic MARS Model

- The output tells us that 32 of 36 terms were used from 25 of the 307 original predictors. If we were to look at all the coefficients, we would see that there are 32 terms in our model (including the intercept). These terms include hinge functions produced from the original 307 predictors (307 predictors because the model uses ‘dummy variables’ for categorical variables).

```
> summary(mars01) %>% .$coefficients %>% head(3)
```

```
                        Sale_Price
(Intercept)           151915.31408
h(Gr_Liv_Area-1194)       55.63663
h(1194-Gr_Liv_Area)      -41.41907
```

```
> summary(mars01) %>% .$coefficients %>% tail(3)
```

```
                              Sale_Price
NeighborhoodNorthridge_Heights  17717.665
Overall_CondGood                 9185.667
Roof_MatlWdShngl                50815.842
```

Fitting a Basic MARS Model

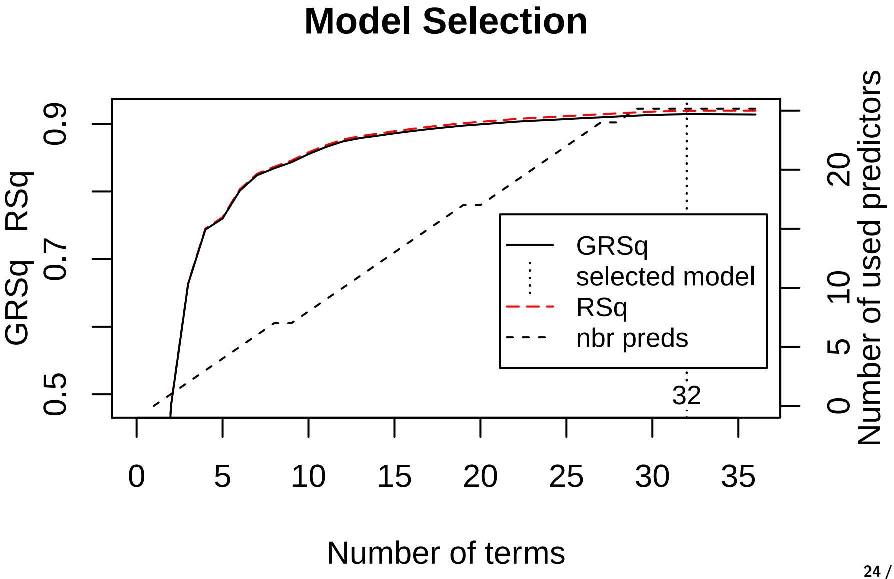
- Looking at the first 3 terms in our model, we see that `Gr_Liv_Area` is included with a knot at 1194. The coefficient for `h(Gr_Liv_Area-1194)` is 55.63663 and the coefficient for `h(1194-Gr_Liv_Area)` is -41.41907.
- In the data set, we see that the `Neighborhood` factor has 28 levels, one for each neighborhood. Looking at the last 3 terms in our model, we see that `Northridge_Heights` has been included in the model with a coefficient of 17717.665. We also see that certain levels of the factors `Overall_Cond` and `Roof_Matl` have also entered the model.

Fitting a Basic MARS Model

- The default plot method for MARS model objects provide a convenient performance plot. The following figure illustrates the model selection plot that graphs the GCV R^2 and R^2 based on the number of terms retained in the model, which are constructed from a certain number of original predictors and factors. The vertical dashed line at 32 tells us the optimal number of terms retained.

Fitting a Basic MARS Model

```
> plot(mars01, which = 1)
```



Extensions and Considerations

- Interactions between different hinge functions
- Tuning the model: the number of retained terms, the degree of interactions
- Variable importance
- Recommended resources:

1. <https://en.wikipedia.org/wiki/Multivariate_adaptive_regression_spline>
2. <http://www.milbo.users.sonic.net/earth/index.html>

WESTERN DATA
ANALYTICS

Thank you for your attention and time!