

# Multivariate Adaptive Regression Splines: A Flexible Approach to Nonlinear Regression

STA 4241 Final Project

Spring 2026

Nick Parmigiano

## 1. Introduction

In many applied regression problems, the relationship between a response variable and its predictors is not linear. Standard linear regression models assume a linear functional form, which can lead to poor predictive performance and misleading inferences when the true relationship involves curves, thresholds, or other nonlinear patterns. While the analyst can attempt to address this by manually adding polynomial terms or applying transformations, this approach requires guessing the correct form in advance, which is impractical when dealing with multiple predictors.

Tree-based methods such as CART (Breiman et al., 1984) offer an alternative by recursively partitioning the predictor space into regions and fitting simple models within each region. Trees handle nonlinearity and interactions automatically, but they produce discontinuous, step-function predictions. The fitted surface jumps abruptly at partition boundaries, which is often unrealistic for continuous outcomes like housing prices or biological measurements.

Multivariate Adaptive Regression Splines (MARS), introduced by Friedman (1991), generalizes earlier work on adaptive spline modeling like the univariate approach of Smith (1982) and the adaptive methods developed in Friedman (1988) to the multivariate setting. MARS provides a flexible, data-driven approach to modeling nonlinear relationships while retaining interpretability. Like tree-based methods, MARS requires no prior specification of the functional form. Like linear regression, the resulting model is a sum of basis functions whose coefficients are estimated by least squares, making it interpretable and computationally efficient.

This paper introduces the MARS algorithm, explains its mathematical foundations, and demonstrates its application to the Ames Housing dataset (De Cock, 2011). We compare MARS against ordinary least squares (OLS) regression and polynomial regression, showing that MARS achieves substantially better predictive accuracy on this dataset due to its ability to capture nonlinear relationships automatically.

## 2. Methodology

### 2.1 Basis Functions

The foundation of MARS is the hinge function, also called a truncated linear basis function. Given a predictor variable  $x$  and a knot location  $t$ , a pair of hinge functions is defined as:

$$h_1(x) = \max(0, x - t)$$

$$h_2(x) = \max(0, t - x)$$

The first function is zero for all values of  $x$  below the knot  $t$ , then increases linearly. The second is linear below  $t$  and zero above it. Together, this reflected pair can represent a change in slope at the knot location.

Unlike polynomial basis functions, which are global (affecting the fitted value across the entire range of  $x$ ), hinge functions are local. Friedman (1991) emphasizes that truncated power basis functions exhibit locality in the sense that modifying a knot location affects only the corresponding basis function, leaving the rest of the basis unchanged. This contrasts with global parametric approaches, which “have limited flexibility and are likely to produce accurate approximations only when the form of the true underlying function is close to the prespecified parametric one”. This locality is a key advantage, as it prevents changes in one region from distorting the fit in distant regions, a problem that plagues high-degree polynomial regression.

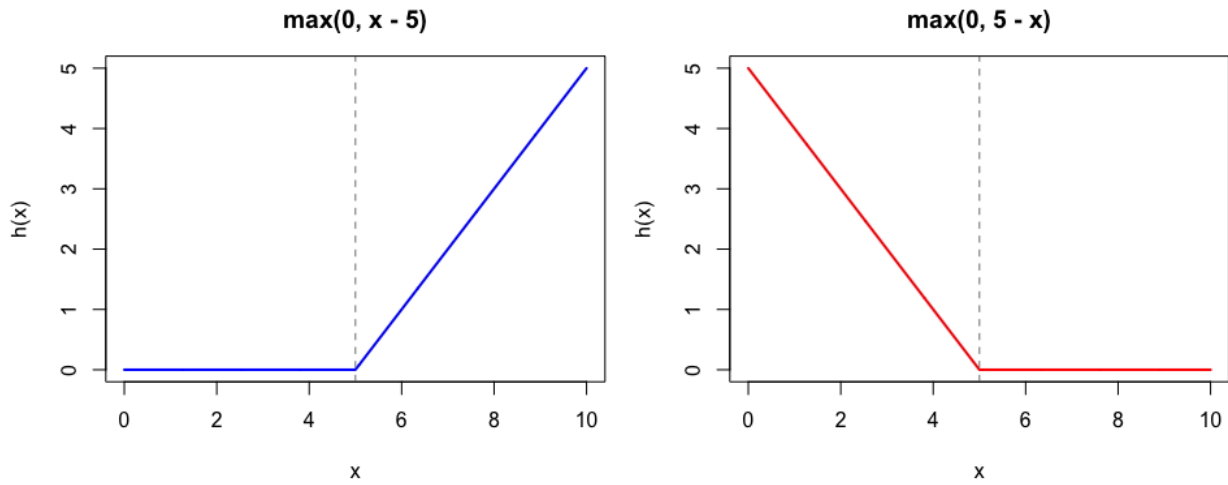


Figure 1. Illustration of hinge (basis) functions used in MARS. Left:  $\max(0, x - 5)$ . Right:  $\max(0, 5 - x)$ .

## 2.2 The MARS Model

A MARS model takes the form of a linear combination of basis functions:

$$f(x) = \beta^0 + \sum_{m=1}^M \beta_m B_m(x)$$

where  $\beta_0$  is the intercept, each  $B_m(x)$  is either a single hinge function or a product of two or more hinge functions (capturing interactions), and the coefficients  $\beta_m$  are estimated by ordinary least squares. The number of basis functions  $M$ , the variables and knot locations involved in each one, and the interaction structure are all determined automatically by the algorithm (Friedman, 1991).

## 2.3 Forward Pass

The MARS algorithm builds the model in two stages. In the forward pass, the algorithm begins with only the intercept term and then greedily adds basis functions to reduce the residual sum of squares (RSS). At each step, the algorithm searches over all predictors and all observed values of each predictor as candidate knot locations. For each candidate, it considers adding the reflected pair of hinge functions and selects the pair that produces the largest decrease in RSS. Specifically, at each step the algorithm solves:

$$\min \text{ over } v, t \text{ of } RSS \text{ after adding } \max(0, x_v - t) \text{ and } \max(0, t - x_v)$$

where  $v$  indexes the predictor variables and  $t$  ranges over the observed values of predictor  $v$ .

Although this search is exhaustive, computational efficiency is achieved through specialized updating formulas for least-squares fitting. These updates avoid recomputing full regressions at each step and make the procedure feasible even for moderate-sized datasets (Friedman, 1991).

The forward pass continues adding pairs of basis functions until a pre-specified maximum number of terms is reached, typically producing an intentionally overfit model. Friedman (1991) explicitly notes that the strategy is to “deliberately overfit the data with an excessively large model and then to trim it back”. This design makes the subsequent backward pass essential, providing a principled mechanism for controlling model complexity rather than relying on early stopping.

## 2.4 Backward Pass and Generalized Cross-Validation

The overfit model from the forward pass is then simplified in the backward pass, which sequentially removes the term whose deletion causes the smallest increase in RSS. This produces a sequence of nested models of decreasing complexity.

To select the best model from this sequence, MARS uses Generalized Cross-Validation (GCV), an approximation to leave-one-out cross-validation introduced by Craven and Wahba (1979) and adapted for MARS by Friedman (1991). The GCV criterion is defined as:

$$GCV = \frac{RSS}{\left[ N \cdot \left( 1 - \frac{\tilde{C}(M)}{N} \right)^2 \right]}$$

where RSS is the residual sum of squares,  $N$  is the number of observations, and  $\tilde{C}(M)$  is a cost-complexity measure. In MARS, this is defined as:

$$\tilde{C}(M) = C(M) + d \cdot M$$

where  $C(M)$  is the trace of the hat matrix, and the term  $d \cdot M$  penalizes the selection of nonlinear basis function parameters during the forward pass (Friedman, 1991). The parameter  $d$  is typically set to 3 for general MARS models and 2 for additive models. Simulation studies in Friedman (1991) show that predictive accuracy is relatively insensitive to the choice of  $d$  within the range  $2 \leq d \leq 4$ , supporting its use as a stable default.

This two-stage procedure is analogous to the grow-then-prune strategy used in CART, but the resulting model is continuous rather than a step function. The GCV criterion avoids the computational expense of performing actual leave-one-out cross-validation while providing a similar bias-variance tradeoff (Hastie et al., 2009).

## 2.5 Comparison with Other Methods

MARS occupies a useful position among regression methods. Compared to linear regression and its regularized variants (ridge, lasso), MARS can capture nonlinear relationships without requiring the analyst to pre-specify polynomial terms or transformations.

Compared to decision trees, MARS produces continuous predictions rather than step functions. Recursive partitioning methods “suffer from some fairly severe restrictions,” most notably that the fitted function is discontinuous at region boundaries, which “severely limits the accuracy of the approximation, especially when the true underlying function is continuous”. By using spline basis functions instead of constant regions, MARS avoids this limitation while retaining the adaptive structure of tree-based methods.

Compared to generalized additive models (GAMs), MARS does not require the analyst to pre-specify which variables or interactions to include. Friedman (1991) notes that additive and interaction spline methods require variable subsets to be selected in advance, whereas MARS automatically selects them based on the data for each problem. This makes MARS particularly useful when the underlying interaction structure is unknown.

### 3. Data Analysis

#### 3.1 Dataset Description

We apply MARS to the Ames Housing dataset, compiled by De Cock (2011) as a modern alternative to the Boston Housing dataset. The dataset describes 2,930 residential property sales in Ames, Iowa from 2006 to 2010, with 80 explanatory variables covering lot characteristics, building quality, square footage, and other features relevant to home valuation. The response variable is the sale price of each home.

For this analysis, we selected 10 numeric predictors that are commonly associated with housing prices: above-grade living area (Gr.Liv.Area), overall quality rating (Overall.Qual), year built (Year.Built), total basement square footage (Total.Bsmt.SF), garage area (Garage.Area), number of full bathrooms (Full.Bath), lot area (Lot.Area), number of fireplaces (Fireplaces), year of most recent remodel (Year.Remod.Add), and first floor square footage (X1st.Flr.SF). After removing observations with missing values, we split the data into a training set (80%,  $n = 2,344$ ) and a test set (20%,  $n = 586$ ) using a fixed random seed for reproducibility.

#### 3.2 Models Compared

We fit four models to the training data and evaluated each on the held-out test set using mean squared error (MSE) and root mean squared error (RMSE):

**Ordinary Least Squares (OLS):** A standard linear regression of sale price on all 10 predictors, assuming linear relationships throughout.

**Polynomial Regression (degree 2):** A linear model with second-degree polynomial terms for four key predictors (living area, overall quality, year built, and total basement square footage) to partially capture nonlinearity, with the remaining predictors entering linearly.

**MARS (degree 1, no interactions):** A MARS model fit using the earth package in R with degree = 1, restricting the model to additive (non-interaction) hinge function terms.

**MARS (degree 2, with interactions):** A MARS model with degree = 2, allowing pairwise interaction terms between hinge functions.

### 3.3 Results

Table 1 presents the test set performance of each model.

**Table 1: Test Set Performance Comparison**

| Model                    | Test MSE      | Test RMSE (\$) |
|--------------------------|---------------|----------------|
| OLS                      | 1,294,291,168 | 35,976         |
| Polynomial (degree 2)    | 1,097,039,749 | 33,122         |
| MARS (no interactions)   | 858,383,934   | 29,298         |
| MARS (with interactions) | 714,847,081   | 26,737         |

The results show a clear progression in predictive accuracy. OLS has the highest test RMSE at approximately \$35,976, meaning its predictions are off by about \$36,000 on average. Adding polynomial terms for four predictors reduces the RMSE to \$33,122, a modest improvement. The additive MARS model (degree 1) substantially outperforms both linear approaches with an RMSE of \$29,298, representing an 18.5% reduction in RMSE compared to OLS. Allowing MARS to include interaction terms (degree 2) further reduces the RMSE to \$26,737, a 25.7% improvement over OLS.

### 3.4 Interpretation of the MARS Model

The additive MARS model (degree 1) selected 17 terms from 8 of the 10 available predictors. Examining the basis functions provides interpretable insights into the housing market. For example, MARS identified a knot at Overall Quality = 7, with the coefficient below this threshold being  $-8,764$  per unit decrease and the coefficient above being  $+42,957$  per unit increase. This indicates that the marginal value of quality accelerates sharply for homes rated above average, consistent with the intuition that luxury finishes command disproportionately higher prices.

Similarly, MARS found a knot at Year Built = 2007. Below this threshold, each additional year of age reduces the sale price by approximately \$359. Above it, the effect reverses dramatically, with each year increasing price by roughly \$9,710. This breakpoint aligns with the 2007–2008 housing crisis, suggesting that homes built after the downturn may reflect post-crisis construction standards or market conditions.

The MARS model with interactions (degree 2) detected five interaction terms. Notable among these is the interaction between Year Built and Total Basement Square Footage, where the effect of basement size on price is amplified for older homes. This makes practical sense: in older homes, a large, finished basement represents a more significant value-add relative to the rest of the property.

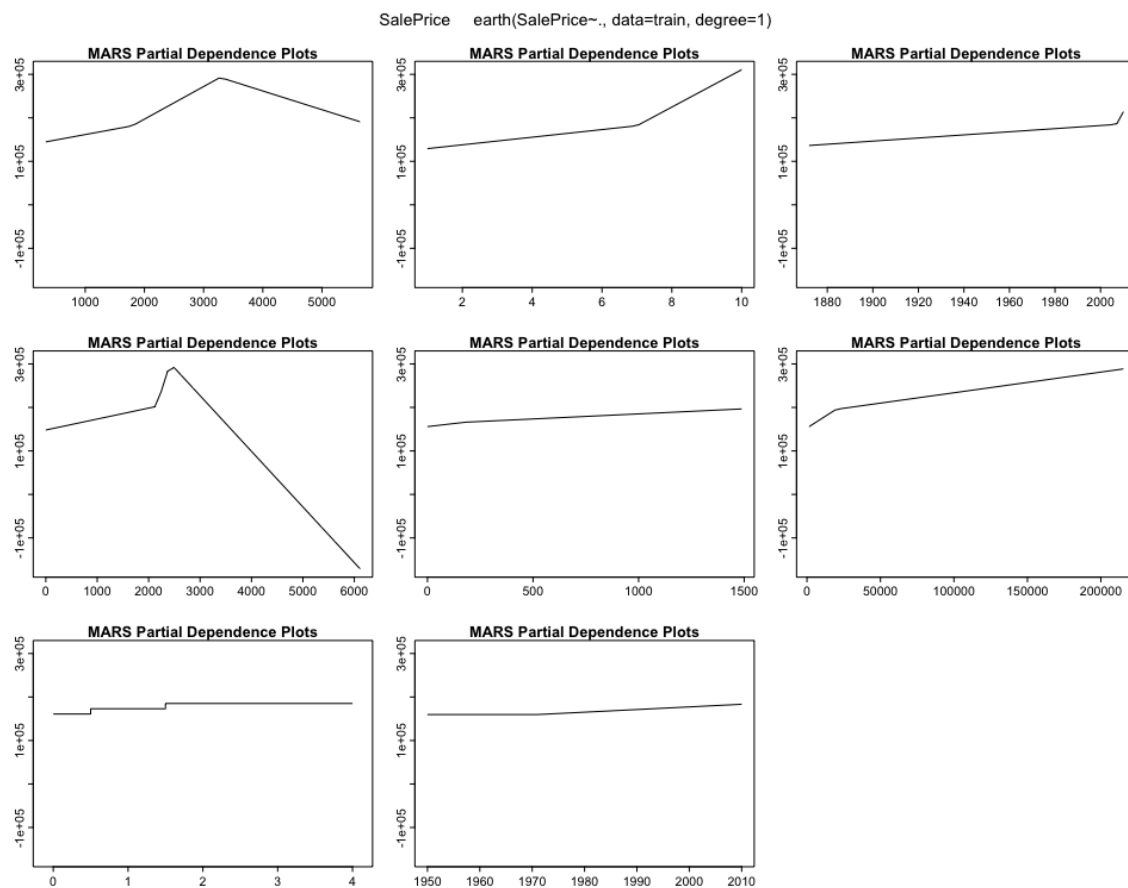


Figure 2. Partial dependence plots from the additive MARS model showing the piecewise linear relationship between each predictor and sale price.

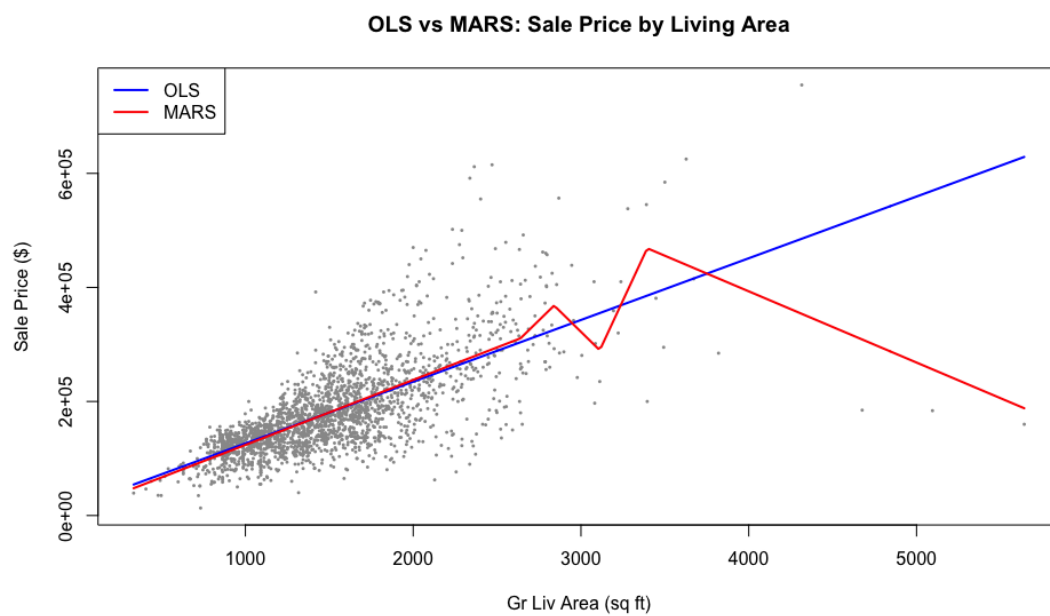


Figure 3. Comparison of OLS (blue) and MARS (red) fitted curves for sale price as a function of above-grade living area.

Figure 3 illustrates the difference between OLS and MARS on a single predictor. The OLS line extends linearly into regions with very few observations, predicting unrealistically high prices for the largest homes. The MARS fit, by contrast, bends at automatically detected knot points, producing a more plausible prediction surface that flattens at the extremes where data is sparse.

### 3.5 Variable Importance

Figure 4 displays the variable importance ranking from the MARS model based on three criteria: the number of model subsets in which each variable appears, and the reduction in GCV and RSS attributable to each variable. Overall Quality is the most important predictor, followed by above-grade living area, year built, and total basement square footage. Lot area, fireplaces, year remodeled, and garage area contribute to the model but are less influential. Full bath count and first floor square footage were not selected by the backward pass, indicating that their predictive contribution was redundant given the other variables

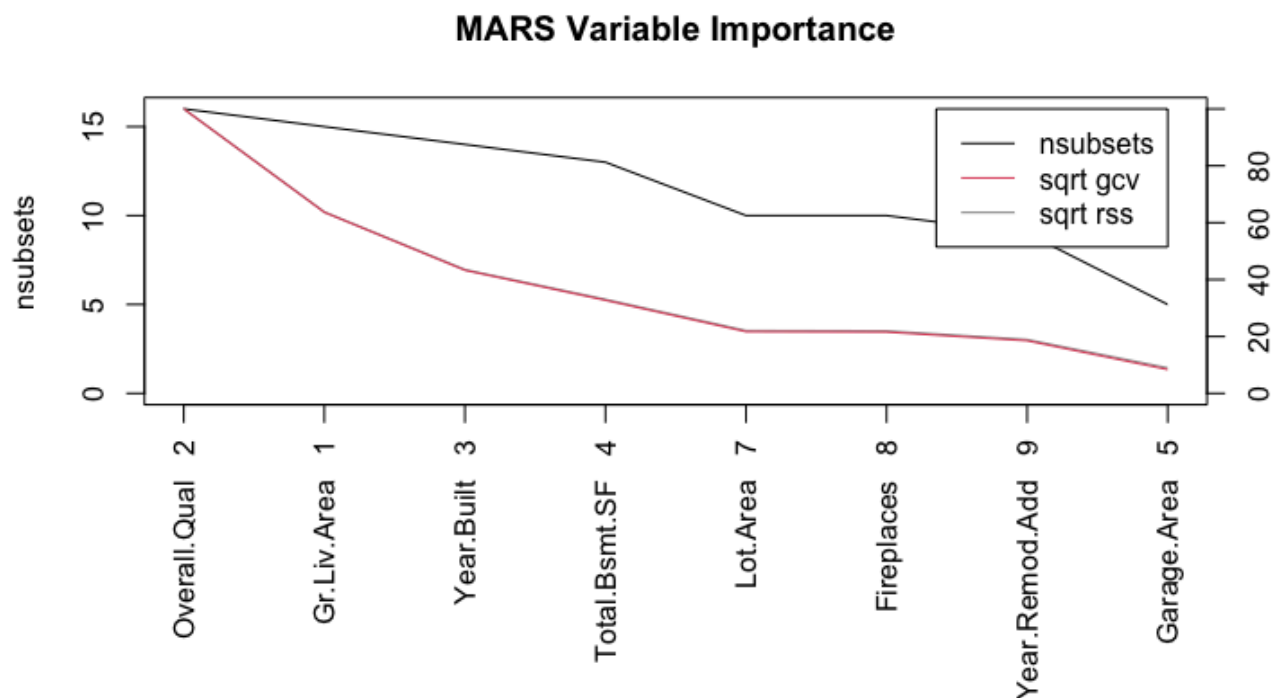
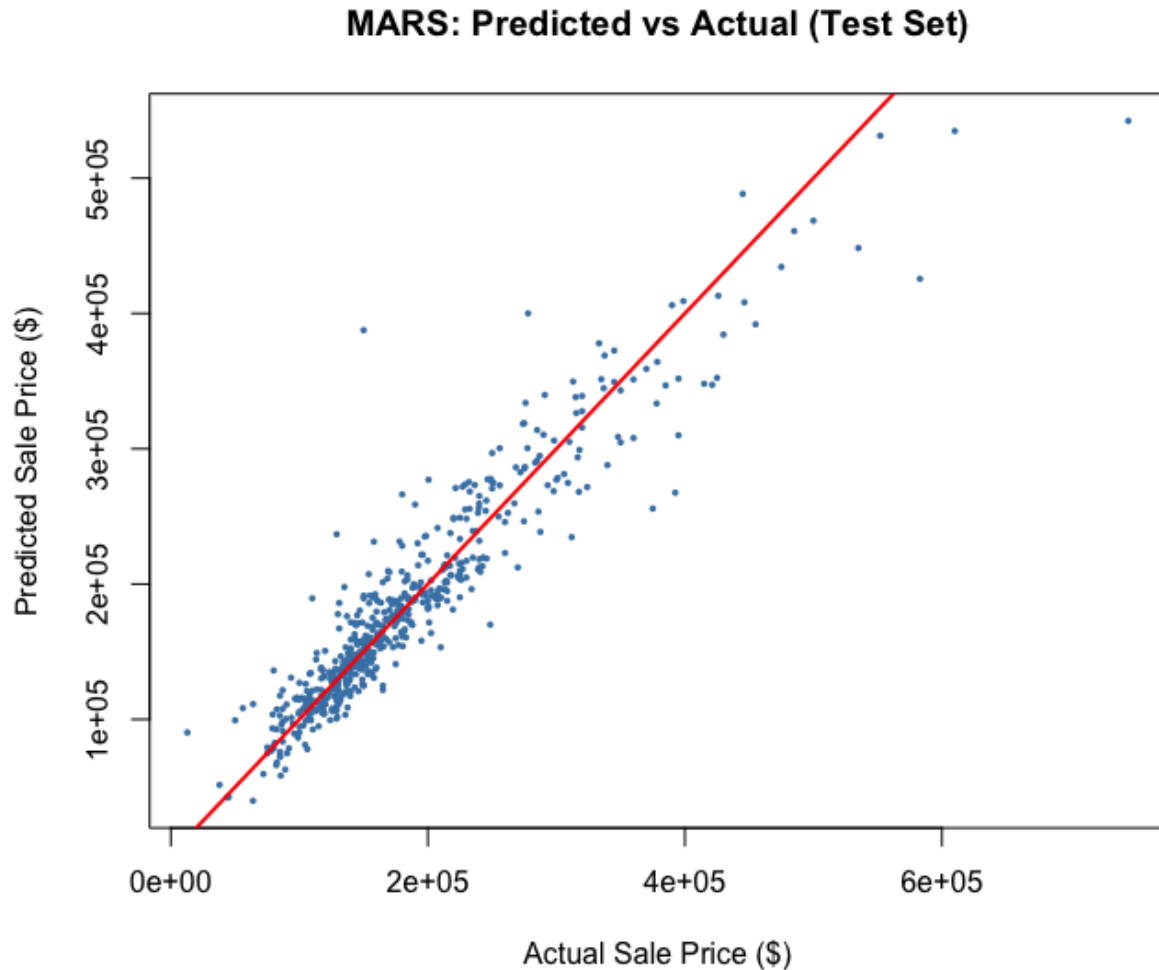


Figure 4. Variable importance from the MARS model ranked by number of subsets, GCV reduction, and RSS reduction.



*Figure 5. Predicted versus actual sale prices on the test set for the MARS model with interactions. The red line indicates perfect prediction.*

Figure 5 shows the predicted versus actual sale prices on the test set for the MARS model with interactions. Points cluster tightly around the 45-degree line for homes priced below \$300,000, with increasing dispersion at higher price points. This pattern is typical in housing price prediction, as luxury homes exhibit more idiosyncratic pricing driven by features not captured in the model.

## 4. Discussion

This paper introduced Multivariate Adaptive Regression Splines (MARS) and demonstrated its application to predicting housing prices. The key findings from our analysis are threefold.

First, MARS substantially outperformed both OLS and polynomial regression in predictive accuracy on the Ames Housing dataset. The additive MARS model reduced test RMSE by 18.5% relative to OLS, and the interaction MARS model achieved a 25.7% reduction. This improvement arises because MARS can identify and model nonlinear relationships automatically, without requiring the analyst to pre-specify the functional form.

Second, the MARS model produces interpretable output. Each basis function corresponds to a specific predictor, knot location, and direction of effect, making it straightforward to understand which variables drive the prediction and where the important thresholds lie. This interpretability distinguishes MARS from black-box methods like neural networks or ensemble methods, where understanding the model's behavior requires additional tools such as partial dependence plots or SHAP values.

Third, MARS occupies a practical middle ground in the bias-variance tradeoff. Linear models have low variance but high bias when the true relationship is nonlinear. Highly flexible methods like random forests or boosted trees have lower bias but higher variance and reduced interpretability. MARS achieves a useful compromise: enough flexibility to capture meaningful nonlinear patterns, but enough structure (piecewise linearity, GCV-based pruning) to avoid severe overfitting.

There are limitations to note. MARS models can be sensitive to correlated predictors; when two features are highly correlated, the algorithm arbitrarily selects whichever it encounters first during the forward pass, potentially making the selected basis functions unstable across different samples. Additionally, while MARS handles moderate-dimensional problems well, it may struggle in very high-dimensional settings ( $p \gg n$ ) where regularized methods like the lasso are more appropriate. Finally, for pure predictive performance on complex datasets, ensemble methods like random forests and gradient boosting typically outperform MARS, though at the cost of interpretability.

Despite these limitations, MARS remains a valuable tool for applied regression problems where the analyst needs both predictive accuracy and interpretability. Its automatic handling of nonlinearity and interactions, combined with the pruning discipline of GCV, makes it a natural first step beyond linear regression for many practical applications.

## References

- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA.
- Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31, 377–403.
- De Cock, D. (2011). Ames, Iowa: Alternative to the Boston Housing Data as an end of semester regression project. *Journal of Statistics Education*, 19(3), 1–15.
- Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1), 1–67.
- Friedman, J. H. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, 4(3), 197–217.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R* (2nd ed.). Springer.

Milborrow, S. (2024). earth: Multivariate Adaptive Regression Splines. R package. Available at <https://CRAN.R-project.org/package=earth>.