Predictive Analytics
Regression and Classification
Lecture 2 : Part 4

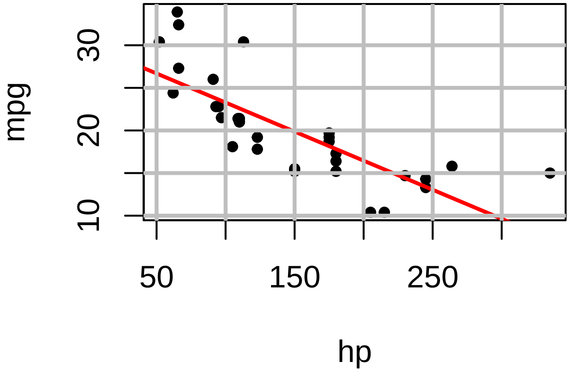
Sourish Das

Chennai Mathematical Institute

Aug-Nov, 2020

Let's understand the Model Complexity

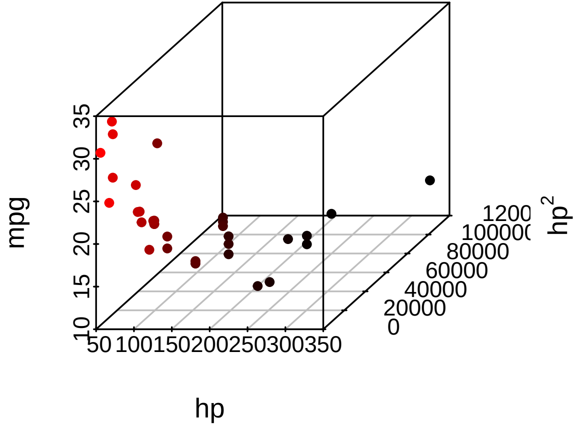
Linear Regression: $\texttt{mpg} = \beta_0 + \beta_1\ \texttt{hp} + \epsilon$



R-squared = 0.602 RMSE = 3.74

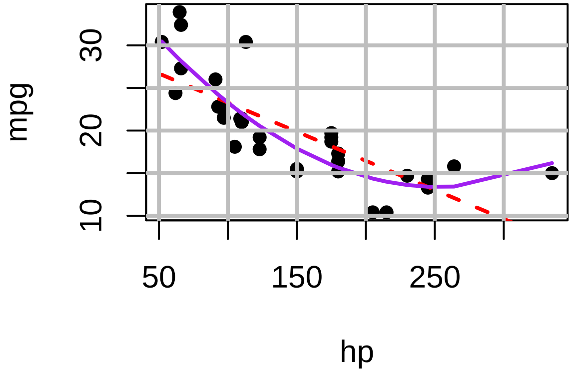
Let's understand the Model Complexity

Quadratic Regression: $\texttt{mpg} = \beta_0 + \beta_1\,\texttt{hp} + \beta_2\,\texttt{hp}^2 + \epsilon$



Let's understand the Model Complexity

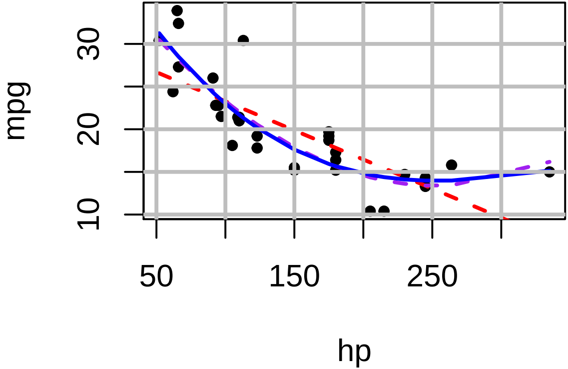
Quadratic Regression: $\texttt{mpg} = \beta_0 + \beta_1\ \texttt{hp} + \beta_2\ \texttt{hp}^2 + \epsilon$



R-squared = 0.756 RMSE = 2.93

Let's understand the Model Complexity

Polynomial Regression of order 3: $\texttt{mpg} = \beta_0 + \beta_1\ \texttt{hp} + \beta_2\ \texttt{hp}^2 + \beta_3\ \texttt{hp}^3 + \epsilon$



R-squared = 0.761 RMSE = 2.903

Let's understand the Model Complexity

```
Model 1     R-squared =  0.602     RMSE =  3.74
Model 2     R-squared =  0.756     RMSE =  2.93
Model 3     R-squared =  0.761     RMSE =  2.903
```

Model Complexity

M1 **Regression Line**
$\texttt{mpg} = \beta_0 + \beta_1 \texttt{ hp} + \epsilon$

M2 **Regression Plane**
$\texttt{mpg} = \beta_0 + \beta_1 \texttt{ hp} + \beta_2 \texttt{ hp}^2 + \epsilon$

M3 **Regression 3-dimension hyper plane**
$\texttt{mpg} = \beta_0 + \beta_1 \texttt{ hp} + \beta_2 \texttt{ hp}^2 + \beta_3 \texttt{ hp}^3 + \epsilon$

M3' **Regression 3-dimension hyper plane**
$\texttt{mpg} = \beta_0 + \beta_1 \texttt{ hp} + \beta_2 \texttt{ wt} + \beta_3 \texttt{ disp} + \epsilon$

Higher order Regression with High Model Complexity

1. Terms for curvature in linear regression

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \epsilon_i$$

   is a polynomial of order p

2. sine cosine functions of incresing frequencies

$$y_i = \beta_1 + \beta_2 \sin(\omega x_i) + \beta_3 \cos(\omega x_i) + \beta_4 \sin(2\omega x_i) + \cdots + \epsilon_i$$

   Such model is also known as Fourier model

Remarks

- The simple LS method can be adopted for any model which is **linear in parameters**.
- **As p the number of predictor or the order of the polynomial increases the model complexity increses.**
- What happens when the model complexity increases too much?
- In order to find the answer for this question; first we try to understand two concepts; (1) Bias and (2) Variance

Bias and Variance

- What is Bias? If θ is and unknown parameter and $\hat{\theta}$ is an estimator of θ, then

$$\mathbb{E}(\hat{\theta}) = \theta + b,$$

where b is known as **bias**. If $b = 0$, then $\hat{\theta}$ is known as **unbiased estimator** of θ

- What is Variance? The variance of $\hat{\theta}$ is

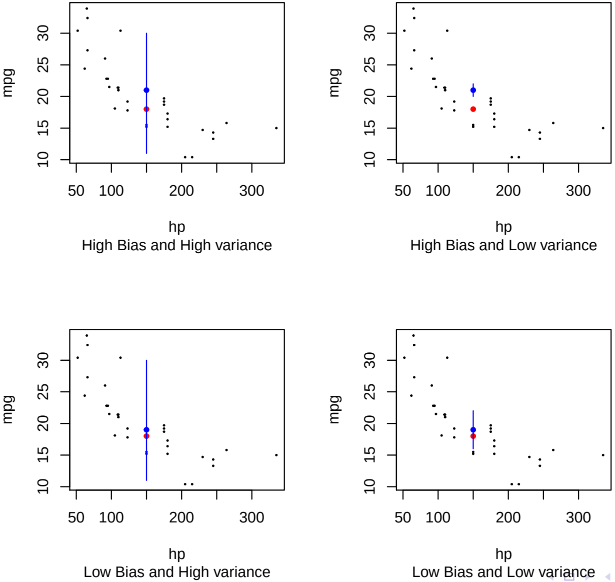
$$\mathbb{V}ar(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2,$$

- Consider the prediction of the new response at input x_0

$$y_0 = f(x_0) + \epsilon_0.$$

$\hat{f}(x_0) = x_0^T \hat{\beta}$ is an estimator of $f(x_0) = x_0^T \beta$.

Bias and Variance



Bias-Variance Tradeoff

- Consider the prediction of the new response at input x_0

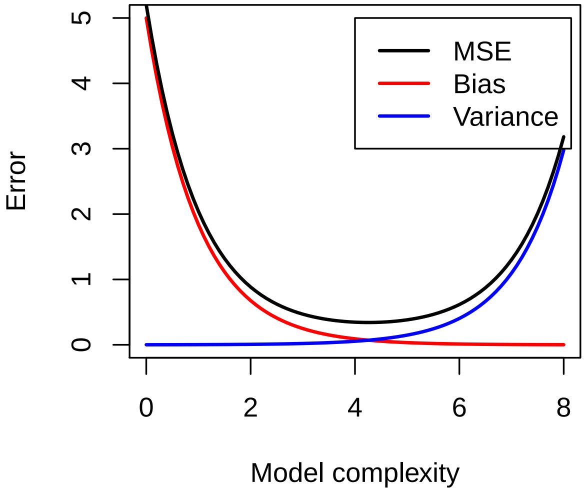
$$y_0 = f(x_0) + \epsilon_0.$$

$\hat{f}(x_0) = x_0^T \hat{\beta}$ is an estimator of $f(x_0) = x_0^T \beta$.

- The MSE of $\hat{f}(x_0) = x_0^T \hat{\beta}$ is

$$\begin{aligned} MSE(\hat{f}(x_0)) &= \mathbb{E}(\hat{y}_0 - f(x_0))^2 \\ &= \mathbb{E}(\hat{y}_0 - \mathbb{E}(\hat{y}_0))^2 + [\mathbb{E}(\hat{y}_0) - f(x_0)]^2 \\ &= \mathbb{V}ar(\hat{y}_0) + Bias^2(\hat{y}_0) \\ &= \mathbb{V}ar(x_0^T \hat{\beta}) + Bias^2(x_0^T \hat{\beta}) \end{aligned}$$

Bias-Variance Tradeoff



In the next part of this lecture...

- We will discuss the technique of variable selection (aka., feature selection) to achieve a parsimonious model with reasonable complexity.