Predictive Analytics
Regression and Classification

Lecture 1 : Part 5

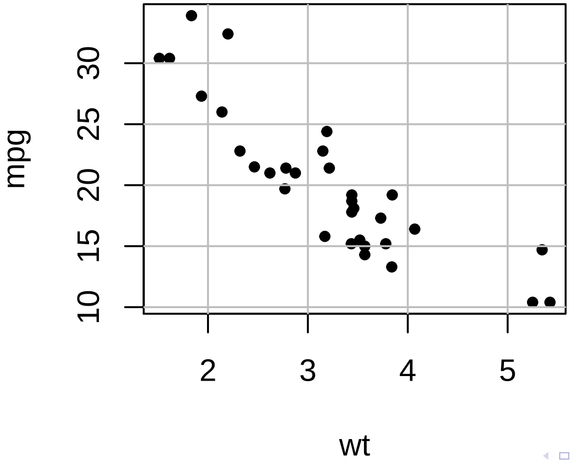
Sourish Das

Chennai Mathematical Institute

Aug-Nov, 2019

Linear Regression

$$\texttt{mpg} = \beta_0 + \beta_1 \ \texttt{wt} + \epsilon$$



Linear Regression

- $\texttt{mpg} = \beta_0 + \beta_1 \ \texttt{wt} + \epsilon$
- We write the model in terms of linear models

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{y} = (\texttt{mpg}_1 \ , \texttt{mpg}_2 \ , \ \ldots \ , \texttt{mpg}_n)^T$;

$$\mathbf{X} = \begin{pmatrix} 1 & \texttt{wt}_1 \\ 1 & \texttt{wt}_2 \\ \vdots & \vdots \\ 1 & \texttt{wt}_n \end{pmatrix}$$

$\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$

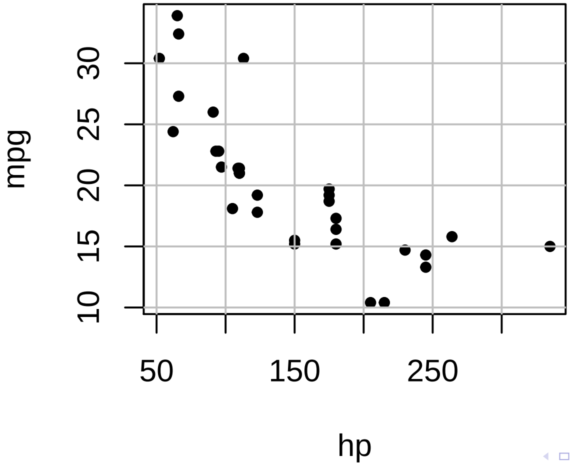
Linear Regression

- **Normal Equations**:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} = (\hat{\beta}_0 \quad \hat{\beta}_1)^T &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= \begin{pmatrix} n & \sum_{i=1}^n \texttt{wt}_i \\ \sum_{i=1}^n \texttt{wt}_i & \sum_{i=1}^n \texttt{wt}_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n \texttt{mpg}_i \\ \sum_{i=1}^n \texttt{wt}_i.\texttt{mpg}_i \end{pmatrix}
\end{aligned}
$$

Quadratic Regression

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



Quadratic Regression

- $\texttt{mpg} = \beta_0 + \beta_1\,\texttt{hp} + \beta_2\,\texttt{hp}^2 + \epsilon$
- We write the model in terms of linear models

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\mathbf{y} = (\texttt{mpg}_1\,, \texttt{mpg}_2\,,\ \ldots\ , \texttt{mpg}_n)^T$;

$$\mathbf{X} = \begin{pmatrix} 1 & \texttt{hp}_1 & \texttt{hp}_1^2 \\ 1 & \texttt{hp}_2 & \texttt{hp}_2^2 \\ \vdots & \vdots & \vdots \\ 1 & \texttt{hp}_n & \texttt{hp}_n^2 \end{pmatrix},$$

$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$ and $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^T$

- The linear model is linear in parameter.

Linear Regression

- **Normal Equations**:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\hat{\beta}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_2)^T \\
&= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= \begin{pmatrix} n & \sum_{i=1}^n \mathrm{hp}_i & \sum_{i=1}^n \mathrm{hp}_i^2 \\ \sum_{i=1}^n \mathrm{hp}_i & \sum_{i=1}^n \mathrm{hp}_i^2 & \sum_{i=1}^n \mathrm{hp}_i^3 \\ \sum_{i=1}^n \mathrm{hp}_i^2 & \sum_{i=1}^n \mathrm{hp}_i^3 & \sum_{i=1}^n \mathrm{hp}_i^4 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n \mathrm{mpg}_i \\ \sum_{i=1}^n \mathrm{hp}_i.\mathrm{mpg}_i \\ \sum_{i=1}^n \mathrm{hp}_i^2.\mathrm{mpg}_i \end{pmatrix}
\end{aligned}
$$

Non-linear Regression Basis Functions

- Consider i^{th} record

  $$y_i = f(\mathbf{x}_i) + \epsilon_i$$

  represents $f(\mathbf{x})$ as

  $$f(\mathbf{x}) = \sum_{j=1}^{K} \beta_j \phi_j(\mathbf{x}) = \boldsymbol{\phi}\boldsymbol{\beta}$$

  we say ϕ is a basis system for $f(\mathbf{x})$.

Representing Functions with Basis Functions

- $\texttt{mpg} = \beta_0 + \beta_1 \texttt{ hp} + \beta_2 \texttt{ hp}^2 + \epsilon$
- Generic terms for curvature in linear regression

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + \cdots + \epsilon_i$$

  implies

$$f(x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \cdots$$

- Sometimes in ML ϕ is known as '**engineered features**' and the process is known as '**feature engineering**'

Fourier Basis

- sine cosine functions of incresing frequencies

$$y = \beta_1 + \beta_2 \sin(\omega x) + \beta_3 \cos(\omega x) + \beta_4 \sin(2\omega x) + \beta_5 \cos(2\omega x) \cdots + \epsilon_i$$

- constant $\omega = 2\pi/P$ defines the period P of oscillation of the first sine/cosine pair. P is known.

- $\boldsymbol{\phi} = \{1, \sin(\omega x), \cos(\omega x), \sin(2\omega x), \cos(2\omega x)...\}$
- $\boldsymbol{\beta}^T = \{\beta_1, \beta_2, \beta_3, \cdots\}$

$$y = \boldsymbol{\phi}\boldsymbol{\beta} + \epsilon$$

- Again in ML $\boldsymbol{\phi}$ is known as '**engineered features**'

- $\texttt{mpg} = \beta_0 + \beta_1 \sin(\omega \; \texttt{hp} \;) + \epsilon$

Functional Estimation/Learning

- We are writing the function with its basis expansion

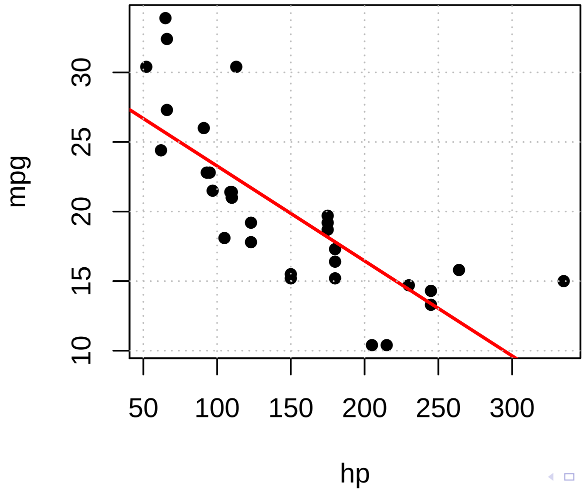
$$y = \phi\boldsymbol{\beta} + \epsilon$$

- Lets assume basis (or **engineered features**) ϕ are **fully known**.

- Problem is $\boldsymbol{\beta}$ is unknown - hence we estimate $\boldsymbol{\beta}$.

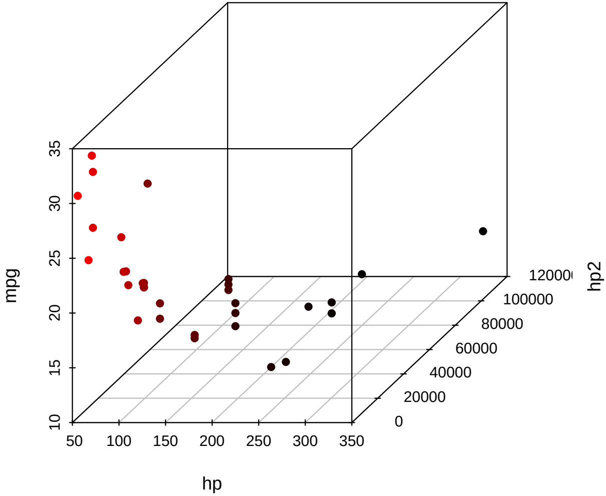
Regression Line

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



Feature Engineering

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



In the next lecture...

- We will discuss sampling distributions and inference of regressions !