Predictive Analytics
Regression and Classification

Lecture 1 : Part 2

Sourish Das

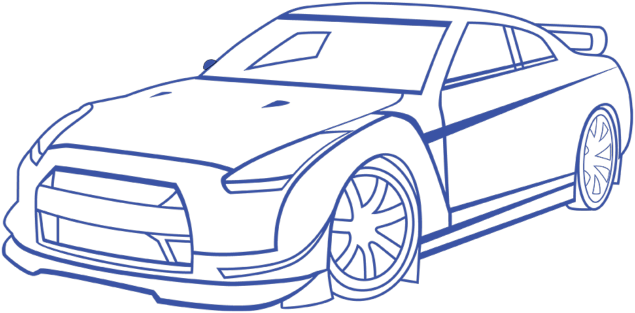
Chennai Mathematical Institute

Aug-Nov, 2019

Regression

Motivating Examples of Linear Regression

Ex 1 Given the different features of a new prototype car, can you predict the mileage or 'miles per gallon' of the car?



Motivating Examples of Regression

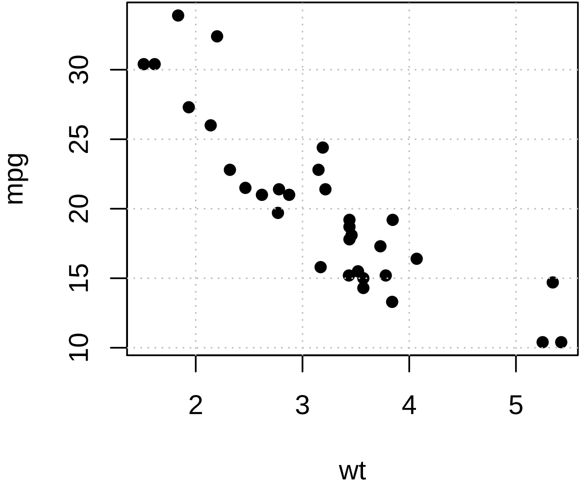
Ex Given the different features of a new prototype car, can you predict the mileage or ‘miles per gallon’ of the car?

	mpg	cyl	disp	hp	wt
Mazda RX4	21.0	6	160	110	2.620
Mazda RX4 Wag	21.0	6	160	110	2.875
Datsun 710	22.8	4	108	93	2.320
Hornet 4 Drive	21.4	6	258	110	3.215
.....					
Prototype	?	4	120	100	3.200

- Note that your objective is to predict the variable `mpg`.

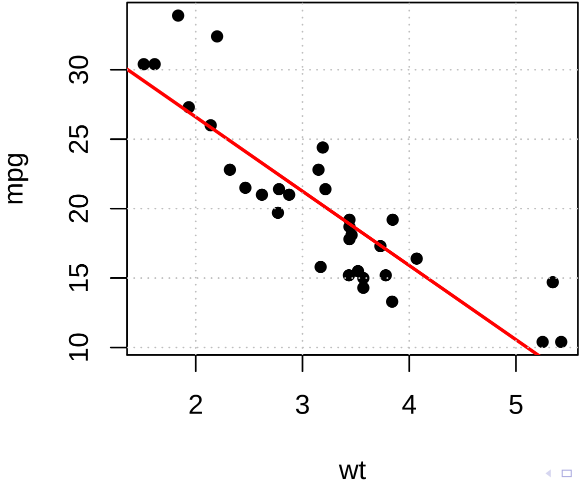
cmi

Plot the data



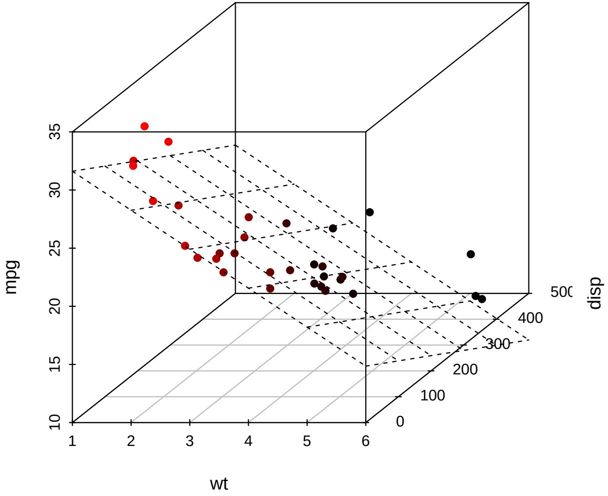
Regression Line

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



Regression Plane

$$\texttt{mpg}=\beta_0+\beta_1\texttt{wt}+\beta_2\ \texttt{disp}+\epsilon$$



Regression Model

- Given a vector of inputs $\mathbf{X}^T = (X_1, X_2, \ldots, X_p)$, we predict the output Y via model

$$Y = \beta_0 + \sum_{j=1}^{p} X_j \beta_j + \epsilon.$$

  The term β_0 is the intercept, also known as the *bias* in machine learning.

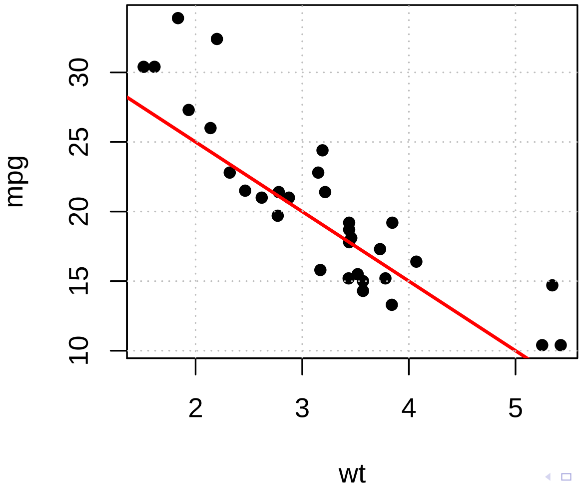
- Often it is convenient to include the constant variable 1 in $\mathbf{X}$, include β_0 in the vector of coefficients $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$

- We have data about y and $\mathbf{X}$

- How can we estimate $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$?

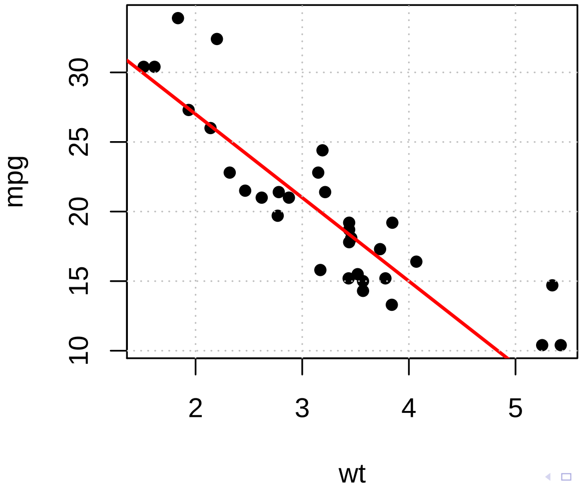
Regression Line

$\texttt{mpg}=35 - 5\texttt{wt}+\epsilon$



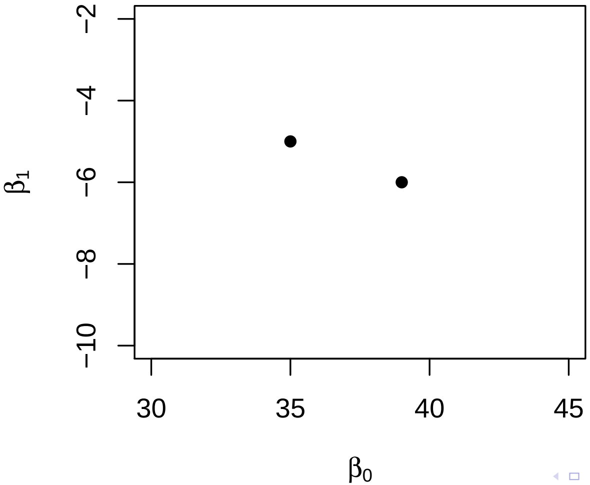
Regression Line

mpg=39 - 6wt+ϵ



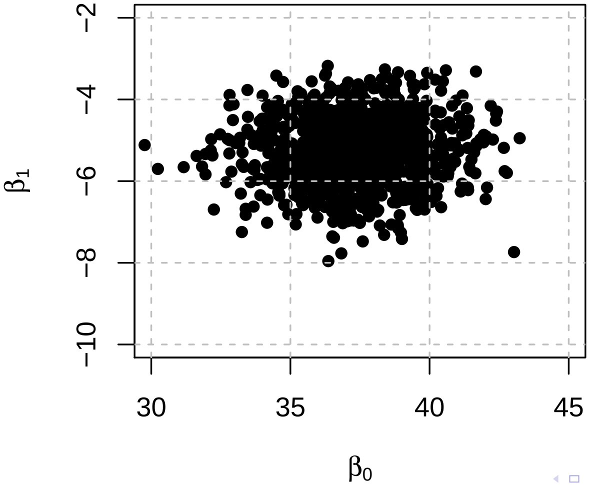
Choice of β

$(\beta_0 = 35,\ \beta_1 = -5)$ and $(\beta_0 = 39,\ \beta_1 = -6)$



Choice of β

However, thousands of choices are there, which one is best?



How do we fit Linear Regression Models?

- Consider the model

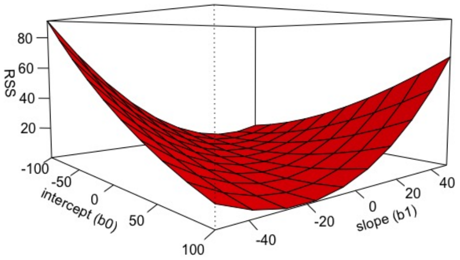
$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}_{n\times 1}$$

- Many different methods, most popular is *least squares.*

- minimize the residual sum of squares

$$\begin{aligned} RSS(\beta) &= (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) \\ &= \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 \end{aligned}$$

Residual Sum of Square : Surface



- $RSS(\beta)$ is a quadratic function of the parameters
- Its minimum always exists, *but may not be unique*.

How do we fit Regression models?

- Differentiate $RSS(\beta)$ with respect to β and equate to 0

$$
\begin{aligned}
& \frac{\partial RSS(\beta)}{\partial \beta} = 0 \\
\implies \quad & \frac{\partial}{\partial \beta}(\mathbf{y} - X\beta)^T(\mathbf{y} - \mathbf{X}\beta) = 0 \\
\implies \quad & -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0 \\
\implies \quad & \mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y} \quad \text{Normal Equations}
\end{aligned}
$$

- $\mathbf{X}^T\mathbf{X}$ is $p \times p$ matrix,
- So normal equations have p unknown and p equations.

In the next part...

- We will discuss how can we solve the normal equations and fit linear regression model....

Thank You

sourish@cmi.ac.in