Predictive Analytics
Regression and Classification

Lecture 1 : Part 3

Sourish Das

Chennai Mathematical Institute

Aug-Nov, 2019

Regression Model

- Given a vector of inputs $\mathbf{X}_{n\times p} = ((X_{ij}))$, we predict the output $\mathbf{y}$ via model

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

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n\times 1}, \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}_{n\times p}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{n\times 1}$$

- It is convenient to include the constant variable 1 in $\mathbf{X}$, to include the intercept.

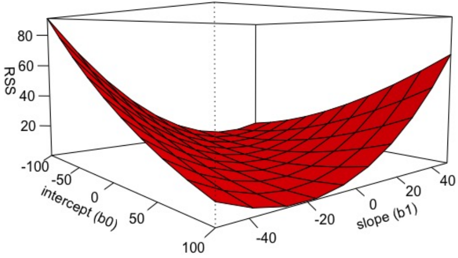
- How can we estimate $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$?

How do we fit Linear Regression Models?

- 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 Linear Regression Models?

- minimize the residual sum of squares

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

$$\frac{\partial}{\partial\beta} RSS(\beta) = 0$$

$$\frac{\partial}{\partial\beta}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta) = 0$$

$$\underbrace{(\mathbf{X}^T\mathbf{X})}_{p\times p}\beta_{p\times 1} = \underbrace{(\mathbf{X}^T\mathbf{y})}_{p\times 1}$$

- p linear equations with p unknowns

System of Equation

- Suppose that for a known matrix $A_{p \times p}$ and vector $b_{p \times 1}$, we wish to find a vector $x_{p \times 1}$ such that

$$Ax = b$$

- The standard approach is ordinary least squares linear regression.

$$\underset{x}{\text{minimize}} \quad ||Ax - b||^2$$

where $||.||$ is the Euclidean norm.

- Solution for x is

$$\hat{x} = A^{-1}b$$

- What happened A is not invertible?

Solution to System of Equation

- If $rank(A|b) > rank(A)$ then solution does not exists.
- If $rank(A|b) = rank(A)$ then at least one solution exists.
- If $rank(A|b) = rank(A) = p$, that is A is a full-rank matrix, then A^{-1} uniquely exists and the solution $\hat{x} = A^{-1}b$ is unique.
- If $rank(A|b) = rank(A) < p$, that is A is a less than full-rank matrix, then x has infinitely many solutions. This is considered as ill-posed problem. Which solution to choose and how to choose?

How do we fit Regression models?

Theorem
For normal equations,

$$rank(\mathbf{X}^T\mathbf{X}|\mathbf{X}^T y) = rank(\mathbf{X}^T\mathbf{X})$$

- Whatever may be your data, irrespective of that, normal equations gurantees at least one solution.
- Atleast one solution always exists - if you adopt least squares method.
- If $\mathbf{X}^T\mathbf{X}$ is nonsingular, i.e., $rank(\mathbf{X}^T\mathbf{X}) = p$, then the unique solution is given by

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y$$

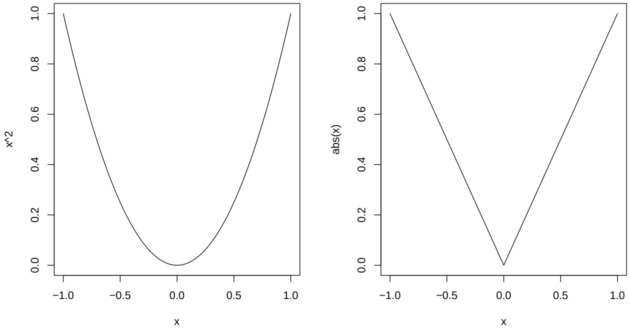
Mean Absolute Deviation?

- What about mean absolute deviation?

$$\Delta(\beta) = \sum_{i=1}^{n} ||y_i - x_i^T \beta||$$

- Conceptually no problem - certainly you can do that.

How do we fit Regression models?



In the next part...

- We will discuss the undelying assumptions of regression models

Thank You

sourish@cmi.ac.in