MM 225 – AI and Data Science

Day 24: Supervised Learning : Regression Analysis-3

Instructors: Hina Gokhale, MP Gururajan, N. Vishwanathan

3 OCTOBER 2024

Outline

Significance of Regression Model

Multiple Regression

Logistic Regression

Method of Gradient Descent

Variety of forms of Regression model

Significance of the regression model

Testing hypothesis that $\beta_1 = 0$ amounts to testing significance of the simple linear model.

Another approach is Analysis of Variance (ANOVA):

Some change in notation:

A and B will be now denoted as $\widehat{\beta_0}$ *and* $\widehat{\beta_1}$ respectively.

Recall:

$SST = Total\ corrected\ sum\ of\ squares = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$

$SSE = Sum\ of\ squares\ of\ residuals = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$

Analysis of Variance Identity

Analysis of variance identity is given by

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Recall:

$SST = Total\ corrected\ sum\ of\ squares = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$

$SSE = Sum\ of\ squares\ of\ residuals = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Define:

$SSR = sum\ of\ squares\ due\ to\ regression = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

Hence symbolically : SST = SSR + SSE

Distributions of SSR and SSE

Recall that

$$SSE = SST - B * Sxy$$

in modified notation : $SSE = SST - \hat{\beta}_1 * Sxy = \mathrm{SST} - \mathrm{SSR}$

Therefore: $SSR = \hat{\beta}_1 * Sxy$

Degrees of freedom for

- SST is (n-1)
- SSE is (n-2)
- SSR is 1

Distributions of SSR and SSE

It can be shown that

$$E\left(\frac{SSE}{n-2}\right) = \sigma^2$$

and

$$E(SSR) = \sigma^2 + \beta_1^2 Sxx$$

Assuming the normality of errors $\epsilon_i, i = 1, 2, \dots, n$

It can be shown that

$$\frac{SSE}{\sigma^2} \sim \chi^2_{n-2}$$

$$\frac{SSR}{\sigma^2} \sim \chi^2_1 \; when \; H_0\colon \; \beta_1 = 0 \; is \; true$$

F Distribution

Let X $\sim \chi^2(n)$ and Y $\sim \chi^2(m)$

Also let X and Y be independent

Then

$$F = \frac{X/n}{Y/m} \sim F(n, m)$$

F is called F distribution

Test for significance of Regression:

When H_0: $\beta_1 = 0 \; is \; true$

$$F_0 = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F(1, n-2)$$

where

MSR = Mean Squares due to regression

MSE = Mean squares due to Error

Hence the null hypothesis is rejected at significance level of α if

$F_0 > \; F(\alpha, 1, n-2)$

ANOVA Table

Source of Variation	**Sums of Squares**	**Degrees of Freedom**	**Mean Squares**	**F - statistic**
Regression	$SSR = \hat{\beta}_1 * Sxy$	1	MSR = SSR/df	MSR/MSE
Error	SSE = SST -SSR	n – 2	MSE = SSE / df	
Total	SST	n – 1		

Multiple Regression

$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$

- where $\beta_0, \beta_1, \ldots \beta_k$ are some constants and ϵ represents random error in this relationship, where $E(\epsilon) = 0 \; and \; Var(\epsilon) = \sigma^2$

Consider n observations $(Y_i, x_{i1}, x_{i2}, \ldots, x_{ik}), i = 1, 2, \ldots, n$

We have system of equations:

$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i \; ; i = 1, 2, \ldots, n$

Need to estimate the regression coefficients $\beta_0, \beta_1, \ldots \beta_k$

properties of these estimators of $\beta_0, \beta_1, \ldots \beta_k$

LSE of Multiple Regression Coefficients

Consider estimation by method of Least Squares Estimation:

Want to find $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ such that sum of squared error from Y_i and the regression expression $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$ is minimized.

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} \sum_{i=1}^{n} \left[Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik}\right)\right]^2$$

Taking partial derivatives wrt $\hat{\beta}_j; j = 0, 1, \ldots, k$ we need to solve following simultaneous linear equations, **called normal equations**

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_{i1} + \hat{\beta}_2 \sum x_{i2} + \cdots + \hat{\beta}_k \sum x_{ik}$$

$$\sum x_{i1} Y_i = \hat{\beta}_0 \sum x_{i1} + \hat{\beta}_1 \sum x_{i1}^2 + \hat{\beta}_2 \sum x_{i1} x_{i2} + \cdots + \hat{\beta}_k \sum x_{i1} x_{ik}$$

$$\vdots$$

$$\sum x_{ik} Y_i = \hat{\beta}_0 \sum x_{ik} + \hat{\beta}_1 \sum x_{ik} x_{i1} + \hat{\beta}_2 \sum x_{ik} x_{i2} + \cdots + \hat{\beta}_k \sum x_{ik}^2$$

Matrix notation:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\ \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots x_{1k} \\ 1 & x_{21} & x_{22} & \cdots x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{and}\ \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$n \times 1$ $\quad n \times (k+1)$ $\quad (k+1) \times 1$ $\quad n \times 1$

Then

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i\, ; i = 1, 2, \ldots, n$$

can be expressed as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

Note that **Y : n x 1; X : n x (k+1) ; $\boldsymbol{\epsilon}$: n x 1** and **$\boldsymbol{\beta}$: (k+1) x 1**

LSE in Matrix Notations

Want to find LSE of $\boldsymbol{\beta}$ by minimizing

$$L = \sum \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

Estimators $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be found by solving following matrix equation

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0$$

the resulting normal equations to be solved in matrix form is given by

$$\boldsymbol{X'X\widehat{\beta}} = \boldsymbol{X'Y}$$

Let us compare with scalar notation…

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_{i1} + \hat{\beta}_2 \sum x_{i2} + \cdots + \hat{\beta}_k \sum x_{ik}$$

$$\sum x_{i1} Y_i = \hat{\beta}_0 \sum x_{i1} + \hat{\beta}_1 \sum x_{i1}^2 + \hat{\beta}_2 \sum x_{i1} x_{i2} + \cdots + \hat{\beta}_k \sum x_{i1} x_{ik}$$

$$\vdots$$

$$\sum x_{ik} Y_i = \hat{\beta}_0 \sum x_{ik} + \hat{\beta}_1 \sum x_{ik} x_{i1} + \hat{\beta}_2 \sum x_{ik} x_{i2} + \cdots + \hat{\beta}_k \sum x_{ik}^2$$

can be expressed as

$$\begin{bmatrix} \sum Y_i \\ \sum x_{i1} Y_i \\ \vdots \\ \sum x_{ik} Y_i \end{bmatrix} = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \dots \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1} x_{i2} & \dots \sum x_{i1} x_{ik} \\ \vdots & \vdots & \vdots & \vdots \\ \sum x_{ik} & \sum x_{ik} x_{i1} & \sum x_{ik} x_{i2} & \dots \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \vdots \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

LSE in Matrix Notations

Want to find LSE of $\boldsymbol{\beta}$ by minimizing

$$L = \sum \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta})$$

Estimators $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ can be found by solving following matrix equation

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0$$

the resulting normal equations to be solved in matrix form is given by

$$\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{Y}$$

which can be solved as

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{X}'\boldsymbol{Y}$$

Properties of LSE

Residuals $\boldsymbol{e} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ where $\widehat{\boldsymbol{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$

In the same lines as in the simple linear regression we have

$$\hat{\sigma}^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-k-1} = \frac{\sum e_i^2}{n-k-1} = \frac{SSE}{n-k-1}$$

Consider $E(\widehat{\boldsymbol{\beta}}) = E\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\right)$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{E}(\boldsymbol{Y}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

Thus $\widehat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$

Properties of LSE

- Similarly it can be shown that, $\boldsymbol{Cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2\boldsymbol{C}$

- Where $\boldsymbol{C} = (\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} C_{00} & \cdots & C_{0k} \\ \vdots & \ddots & \vdots \\ C_{k0} & \cdots & C_{kk} \end{bmatrix}$ then

$$Var(\hat{\beta}_j) = \sigma^2 C_{jj} \quad for\ j = 0, 1, \ldots, k$$

$$cov(\hat{\beta}_j, \hat{\beta}_l) = \sigma^2 C_{jl} \ \ for\ j \neq l\ ; j, l = 0, 1, \ldots, k$$

Hypothesis testing in Multiple Regression

Consider H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

The alternate hypothesis is H_A: $\beta_j \neq 0$ $for\ at\ least\ one\ j$

- This implies that at least one x_j : j = 1, 2, ... , k contributes significantly to the model.

Procedure remains same as in simple linear regression:

The ANOVA identity holds here as

$$SST = SSR + SSE$$

$E(SSR) = \sigma^2 + \beta'(X'X)^{-1}\beta$

But when H_0 is true: $\frac{SSR}{\sigma^2} \sim \chi_k^2$

Also, similar to the simple linear model $\frac{SSE}{\sigma^2} \sim \chi_{n-k-1}^2$

Hypothesis testing in Multiple Regression

Consider H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

The alternate hypothesis is H_A: $\beta_j \neq 0$ $for\ at\ least\ one\ j$

- This implies that at least one x_j : j = 1, 2, … , k contributes significantly to the model.

Procedure remains same as in simple linear regression:

The ANOVA identity holds here as

$$SST = SSR + SSE$$

$E(SSR) = \sigma^2 + \beta'(X'X)^{-1}\beta$

But when H_0 is true: $\frac{SSR}{\sigma^2} \sim \chi^2_k$

Also, similar to the simple linear model $\frac{SSE}{\sigma^2} \sim \chi^2_{n-k-1}$

ANOVA Table for multiple regression

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares	F - statistic
Regressio n	SSR = SST - SSE	k	MSR = SSR/k	MSR/MSE
Error	SSE = $\sum(y_i - \hat{y}_i)^2$	n – k – 1	MSE = SSE / n-k-1	
Total	SST = $\sum(y_i - \bar{y}_i)^2$	n – 1		

Large F values indicate deviation from the null hypothesis

Coefficient of Determination

Definition of Coefficient of Determination R^2 remains same in the case of multiple regression model

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

This measure R^2 has a problem: with every additional independent variable R^2 tends to increase.

To overcome this ***Adjusted R²*** is used:

$$R^2_{adj} = 1 - \frac{SSE/n-k-1}{SST/n-1}$$

This way R^2_{adj} will increase with additional variable only if MSE is reduced.

Summary

Linear Regression model to express simple causal relationship between RV Y and r independent variables.

Regression model as model to explore empirical relationship.

Errors are random, with mean 0 and common variance.

Assumptions are sufficient to estimate the regression parameters, their SE and common variance

To test the hypothesis that regression parameter take on particular value additional assumption is made: Errors are normally distributed.