Ch 3.2: Multiple Linear Regression

Lecture 6 - CMSE 381

Prof. Mengsen Zhang

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Mon, Jan 27, 2025

Announcements

Last time:

- 3.1 (Simple) linear regression

Announcements:

- Homework #2 Due Sunday on Crowdmark

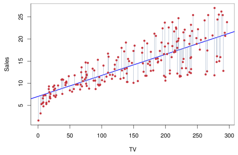
Covered in this lecture

- Multiple linear regression
- Hypothesis test in that case
- Forward and Backward Selection

Section 1

Review from last time

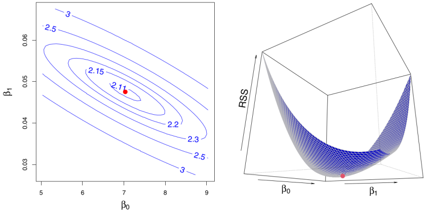
Linear Regression with One Variable



- Predict Y on a single variable X

$$Y \approx \beta_0 + \beta_1 X$$

- Find good guesses for $\hat{\beta}_0$, $\hat{\beta}_1$.
- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- $e_i = y_i - \hat{y}_i$ is the ith residual
- residual sum of squares $\text{RSS} = \sum_i e_i^2$



- RSS is minimized at *least squares coefficient estimates*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Evaluating the model

- Linear regression is unbiased
- Variance of linear regression estimates:

$$\mathrm{SE}(\hat{\beta}_0) = \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \right]$$

$$\mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

where $\sigma^2 = \mathrm{Var}(\varepsilon)$

- Estimate σ: $\hat{\sigma}^2 = \frac{RSS}{n-2}$

- The 95% confidence interval for β_1 approximately takes the form

$$\hat{\beta}_1 \pm 2 \cdot \mathrm{SE}(\hat{\beta}_1)$$

- Hypothesis test:

  H_0: $\beta_1 = 0$
  H_a: $\beta_1 \neq 0$
  - Test statistic $t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}$

Assessing the accuracy of the model

Residual standard error (RSE):

$$RSE = \sqrt{\frac{1}{n-2} RSS}$$

R squared:

$$R^2 = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS}$$

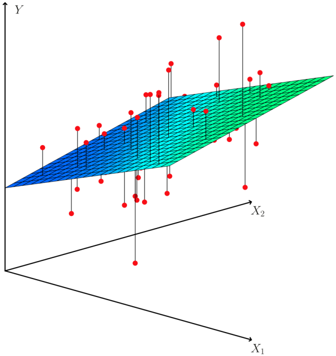
$$TSS = \sum_i (y_i - \overline{y})^2$$

Section 2

Multiple Linear Regression

Setup

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_p X_p + \varepsilon$$



Estimation and Prediction

Given estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_p$, prediction is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_p x_p$$

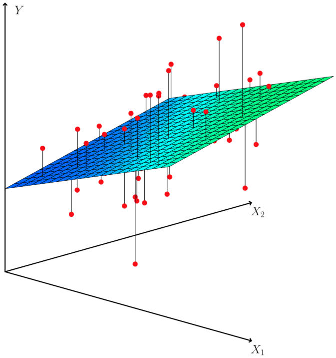
Minimize the sum of squares

$$\begin{aligned} RSS &= \sum_i (y_i - \hat{y}_i)^2 \\ &= \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip})^2 \end{aligned}$$

Coefficients are closed form but UGLY

Advertising data set example

$$\texttt{Sales} = \beta_0 + \beta_1 \cdot \texttt{TV} + \beta_2 \cdot \texttt{radio} + \beta_3 \cdot \texttt{newspaper}$$



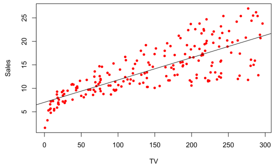
	Coefficient
Intercept	2.939
TV	0.046
radio	0.189
newspaper	−0.001

Interpretation of coefficients

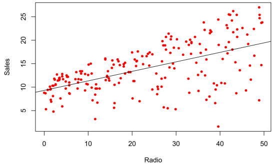
$$\texttt{Sales} = \beta_0 + \beta_1 \cdot \texttt{TV} + \beta_2 \cdot \texttt{radio} + \beta_3 \cdot \texttt{newspaper}$$

	Coefficient
Intercept	2.939
TV	0.046
radio	0.189
newspaper	-0.001

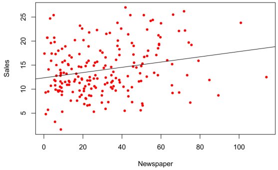
Single regression vs multi-regression



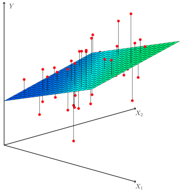
	Coefficient
`Intercept`	7.0325
`TV`	0.0475



	Coefficient
`Intercept`	9.312
`radio`	0.203



	Coefficient
`Intercept`	12.351
`newspaper`	0.055



	Coefficient
`Intercept`	2.939
`TV`	0.046
`radio`	0.189
`newspaper`	−0.001

Correlation matrix

	`TV`	`radio`	`newspaper`	`sales`
`TV`	1.0000	0.0548	0.0567	0.7822
`radio`		1.0000	0.3541	0.5762
`newspaper`			1.0000	0.2283
`sales`				1.0000

Coding group work

Run the section titled "Multiple Linear Regression"

Section 3

Ch 3.2.2: Questions to ask of your regression

Question 1

Is at least one of the predictors $X_1, \cdots, X_p$ useful in predicting the response?

Q1: Hypothesis test

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_p = 0$

H_a : At least one β_j is non-zero

F-statistic:

$$F = \frac{(TSS - RSS)/p}{RSS/(n-p-1)} \sim F_{p,n-p-1}$$

The F-statistic for the hypothesis test

$$F = \frac{(TSS - RSS)/p}{RSS/(n-p-1)} \sim F_{p,n-p-1}$$

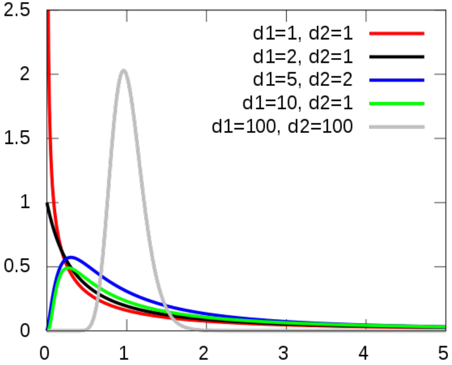


Image from wikipedia, By IkamusumeFan - Own work, CC BY-SA 4.0,

Do Q1 section in jupyter notebook

Q2

Do all the predictors help to explain Y, or is only a subset of the predictors useful?

Q2: A first idea

Great, you know at least one variable is important, so which is it?....

Do Q2 section in jupyter notebook

Why is this a bad idea?

Next time

CMSE381_S2025_Schedule : Sheet1

Lec #	Date		Topic	Reading	HW
1	M	1/13	Intro / Python Review	1	
2	W	1/15	What is statistical learning	2.1	
3	F	1/17	Assessing Model Accuracy	2.2.1, 2.2.2	
	M	1/20	MLK - No Class		
4	W	1/22	Linear Regression	3.1	
5	F	1/24	More Linear Regression	3.1	HW #1 Due Sun 1/26
6	M	1/27	Multi-linear Regression	3.2	
7	W	1/29	Probably More Linear Regression	3.3	
8	F	1/31	Last of the Linear Regression		HW #2 Due Sun 2/1
9	M	2/3	Intro to classification, Bayes classifier, KNN classifier	2.2.3	
10	W	2/5	Logistic Regression	4.1, 4.2, 4.3.1-3	
11	F	2/7	Multiple Logistic Regression / Multinomial Logistic Regression	4.3.4-5	HW #3 Due Sun 2/9
	M	2/10	***Project Day & Review***		
	W	2/12	**Midterm #1**		
12	F	2/14	Leave one out CV	5.1.1, 5.1.2	
13	M	2/17	k-fold CV	5.1.3	
14	W	2/19	More k-fold CV	5.1.4-5	
15	F	2/21	k-fold CV for classification	5.1.5	HW #4 Due Sun 2/23
16	M	2/24	Subset selection	6.1	