Ch 3.2.2: Questions to ask of Multiple Linear Regression

Lecture 6 - CMSE 381

Prof. Elizabeth Munch

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

Weds, Sep 14, 2022

Announcements

Last time:

- 3.1 Linear regression
- Started 3.2 Multiple Linear regression

Announcements:

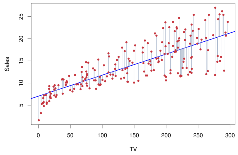
- Homework #2 Due TONIGHT on Crowdmark

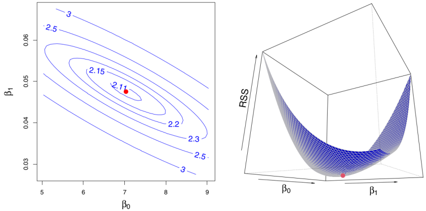
Covered in this lecture

- Multiple linear regression
- Hypothesis test in that case
- Forward and Backward Selection
- R^2 and RSE
- Confidence intervals and prediction intervals

Section 1

Review from last time





- 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
- 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:

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

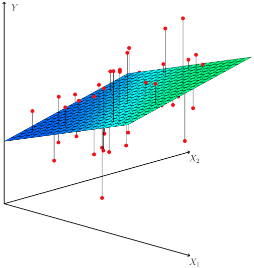
$$TSS = \sum_i (y_i - \overline{y})^2$$

Least Squares Prediction for Multiple Linear Regression

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

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

$$\begin{aligned} RSS &= \sum_i (y_i - \hat{y}_i)^2 \\ &= \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \cdots - \hat{\beta}_p x_p) \end{aligned}$$

- Coefficients are closed form but UGLY
- β_j is average effect on Y for one unit increase in X_j if all other X_i stay fixed

Section 2

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

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

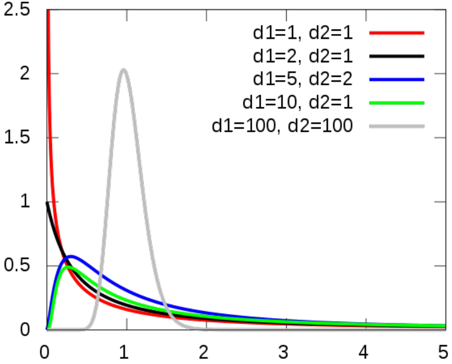


Image from wikipedia, By IkamusumeFan - Own work, CC BY-SA 4.0,

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?

Q3

How well does the model fit the data?

Assessing the accuracy of the module

Almost the same as before

Residual standard error (RSE):

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

R squared:

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

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

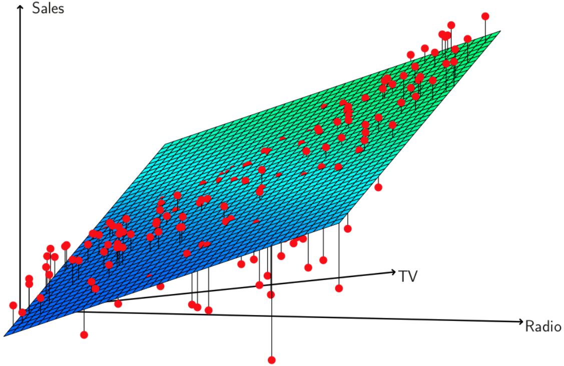
R^2 on Advertising data

- Just TV: $R^2 = 0.61$
- Just TV and radio: $R^2 = 0.89719$
- All three variables: $R^2 = 0.8972$

RSE on Advertising Data

- Just TV: $RSE = 3.26$
- Just TV and radio: $RSE = 1.681$
- All three variables: $RSE = 1.686$

If all else fails, look at the data



Q4

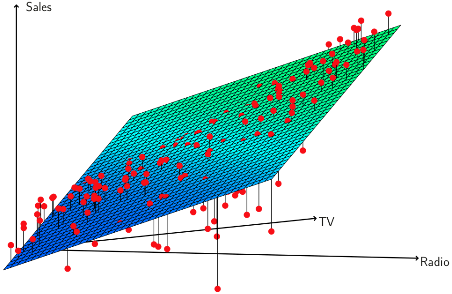
Given a set of predictor values, what response value should we predict, and how accurate is our prediction?

Given estimates $\hat{\beta}_0, \cdots, \hat{\beta}_p$ for $\beta_0, \cdots, \beta_p$
Least squares plane:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_p X_p$$

estimate for the true population regression plane

$$f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$



Confidence intervals and Prediction Intervals

Confidence Interval

The range likely to contain the population parameter (mean, standard deviation) of interest.

Prediction Interval

The range that likely contains the value of the dependent variable for a single new observation given specific values of the independent variables.

Specific to the Advertising Data

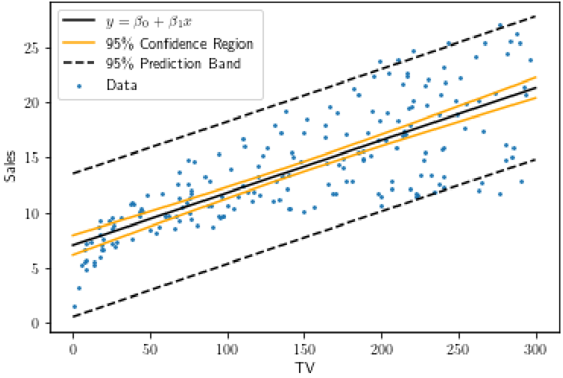
Compute a confidence interval to determine how close $\hat{Y}$ will be to $f(X)$

Advertising example:
If \$100K is spent on TV, and \$20K on radio, the 95% CI for sales is [10,985, 11,528].

Compute a prediction interval to quantify the uncertainty in sales for a particular city.

Advertising example:
Given that \$100,000 is spent on TV advertising and \$20,000 is spent on radio advertising in that city the 95% prediction interval is [7,930, 14,580].

Comparing the two



Go take a look at the code under Q4

Next time

Lec #		Date	Topic	Reading	Homeworks
1	W	Aug 31	Intro / First day stuff / Python Review Pt 1	1	
2	F	Sep 2	What is statistical learning?	2.1	
	M	Sep 5	No class - Labor day		
3	W	Sep 7	Assessing Model Accuracy	2.2.1, 2.2.2	HW #1 Due
4	F	Sep 9	Linear Regression	3.1	
5	M	Sep 12	More Linear Regression	3.1/3.2	
6	W	Sep 14	Even more linear regression	3.2.2	HW #2 Due
7	F	Sep 16	Probably more linear regression	3.3	
8	M	Sep 19	Intro to classification, Logisitic Regression	2.2.3, 4.1, 4.2, 4.3	
9	W	Sep 21	More logistic regression		HW #3 Due
10	F	Sep 23	Review		
11	M	Sep 26	***Midterm #1***		
12	W	Sep 28	[No class, Dr Munch out of town]		
13	F	Sep 30	[No class, Dr Munch out of town]		
14	M	Oct 3	Leave one out CV	5.1.1, 5.1.2	
15	W	Oct 5	k-fold CV	5.1.3	
16	F	Oct 7	More k-fold CV	5.1.4	
17	M	Oct 10	CV for classification	5.1.5	HW #4 Due
18	W	Oct 12	Resampling methods: Bootstrap	5.2	
19	F	Oct 14	Subset selection	6.1	
20	M	Oct 17	Shrinkage: Ridge	6.2.1	HW #5 Due
21	W	Oct 19	Shrinkage: Lasso	6.2.2	
22	F	Oct 21	Dimension Reduction	6.3	