Ch 3.3: The Last of the Linear Regression

Lecture 8 - CMSE 381

Prof. Mengsen Zhang

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

Fri, Jan 31, 2025

Announcements

Last time:

- Started 3.3 Questions of linear regression

Announcements:

- HW #2 Due Sunday 2/1

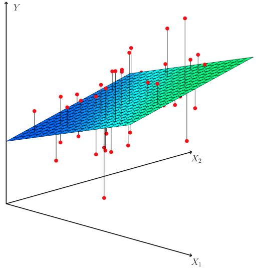
Covered in this lecture

- Extending the linear model with interaction terms
- Hierarchy principle
- Polynomial regression

Section 1

Review from last time

Linear Regression with Multiple Variables



- Predict Y on a multiple variables X

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

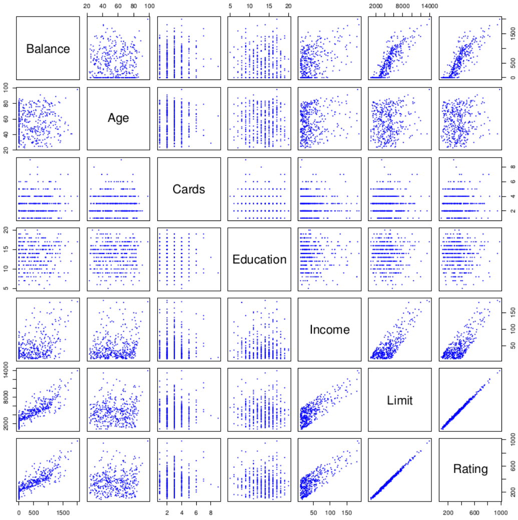
- Find good guesses for $\hat{\beta}_0$, $\hat{\beta}_1, \cdots$.
- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \cdots + \hat{\beta}_p x_p$

- $e_i = y_i - \hat{y}_i$ is the ith residual
- $\text{RSS} = \sum_i e_i^2$
- RSS is minimized at *least squares coefficient estimates*

Section 2

Categorical Input Variables

Credit card balance



- `own`: house ownership
- `student`: student status
- `status`: marital status
- `region`: East, West, or South

One-hot encoding of categorical input variable

Create a new variable

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is a student} \\ 0 & \text{if } i\text{th person is not a student} \end{cases}$$

Model:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \\ &= \begin{cases} \beta_0 + \beta_1 + \varepsilon_i & \text{if } i\text{th person is student} \\ \beta_0 + \varepsilon_i & \text{if } i\text{th person isn't} \end{cases} \end{aligned}$$

Who cares about 0/1?

Old version: 0/1

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is a student} \\ 0 & \text{if } i\text{th person is not a student} \end{cases}$$

Model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$= \begin{cases} \beta_0 + \beta_1 + \varepsilon_i & \text{if } i\text{th person is student} \\ \beta_0 + \varepsilon_i & \text{if } i\text{th person isn't} \end{cases}$$

Alternative version: ± 1

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is a student} \\ -1 & \text{if } i\text{th person is not a student} \end{cases}$$

Model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
$$= \begin{cases} \beta_0 + \beta_1 + \varepsilon_i & \text{if } i\text{th person is student} \\ \beta_0 - \beta_1 + \varepsilon_i & \text{if } i\text{th person isn't} \end{cases}$$

$\hat{y}_i$

Qualitiative Predictor with More than Two Levels

Region:

	x_{i1}	x_{i2}
South		
West		
East		

Create spare dummy variables:

$$x_{i1} = \begin{cases} 1 & \text{if } i\text{th person from South} \\ 0 & \text{if } i\text{th person not from South} \end{cases}$$

$$x_{i2} = \begin{cases} 1 & \text{if } i\text{th person from West} \\ 0 & \text{if } i\text{th person not from West} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

$$= \begin{cases} \beta_0 + \beta_1 + \varepsilon_i & \text{if } i\text{th person from South} \\ \beta_0 + \beta_2 + \varepsilon_i & \text{if } i\text{th person from West} \\ \beta_0 + \varepsilon_i & \text{if } i\text{th person from East} \end{cases}$$

More on multiple levels

	Coefficient	Std. error	t-statistic	p-value
Intercept	531.00	46.32	11.464	< 0.0001
region[South]	-18.69	65.02	-0.287	0.7740
region[West]	-12.50	56.68	-0.221	0.8260

South - east

West - east

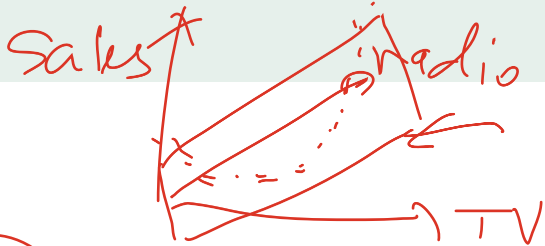
Do code section on "Dummy Variables for Multi-level Categorical Inputs"

Section 3

Extending the linear model

Assumptions so far

Back to our Advertising data set



$$\hat{Y}_{sales} = \beta_0 + \beta_1 \cdot X_{TV} + \beta_2 \cdot X_{radio} + \beta_3 \cdot X_{newspaper}$$

Assumed (implicitly) that the effect on sales by increasing one medium is independent of the others.

What if spending money on radio advertising increases the effectiveness of TV advertising? How do we model it?

Interaction Term

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

$$\begin{aligned}
Y_{sales} &= \beta_0 + \beta_1 X_{TV} + \beta_2 X_{radio} + \beta_3 X_{radio} X_{TV} + \varepsilon \\
&= \beta_0 + (\beta_1 + \beta_3 X_{radio}) X_{TV} + \beta_2 X_{radio} + \varepsilon \\
&= \beta_0 + \tilde{\beta}_1 X_{TV} + \beta_2 X_{radio} + \varepsilon
\end{aligned}$$

main effects

radio x TV

interaction effect

nonlinear

Interaction term

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

	Coefficient	Std. error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

$$\begin{aligned} Y_{sales} &= \beta_0 + \beta_1 X_{TV} + \beta_2 X_{radio} + \beta_3 X_{radio} X_{TV} + \varepsilon \\ &= \beta_0 + (\beta_1 + \beta_3 X_{radio}) X_{TV} + \beta_2 X_{radio} + \varepsilon \end{aligned}$$

Interpretation

	Coefficient	Std. error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

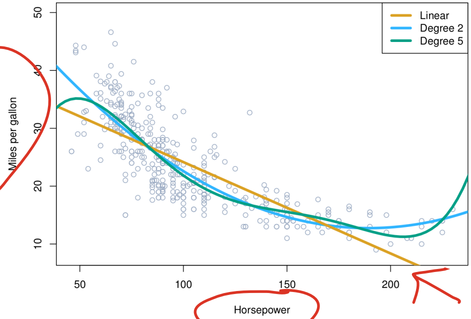
Hierarchy principle

Sometimes p-value for interaction term is very small, but associated main effects are not.

The hierarchy principle:

Nonlinear relationships

$$\texttt{mpg} = \beta_0 + \beta_1 \cdot \texttt{horsepower} + \beta_2 \cdot \texttt{horsepower}^2 + \varepsilon$$



	Coefficient	Std. error	t-statistic	p-value
Intercept	56.9001	1.8004	31.6	< 0.0001
horsepower	−0.4662	0.0311	−15.0	< 0.0001
horsepower^2	0.0012	0.0001	10.1	< 0.0001

Do the section on “Interaction Terms”

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	