

# Math 610:Homework 2

Neal Kuperman

January 15, 2026

## ISLP 3.10

This question should be answered using the Carseats data set.

- (a) Fit a multiple regression model to predict Sales using Price, Urban, and US.
- (b) Provide an interpretation of each coefficient in the model. Be careful—some of the variables in the model are qualitative!
- (c) Write out the model in equation form, being careful to handle the qualitative variables properly.
- (d) For which of the predictors can you reject the null hypothesis  $H_0 : \beta_j = 0$ ?
- (e) On the basis of your response to the previous question, fit a smaller model that only uses the predictors for which there is evidence of association with the outcome.
- (f) How well do the models in (a) and (e) fit the data?
- (g) Using the model from (e), obtain 95% confidence intervals for the coefficient(s).
- (h) Is there evidence of outliers or high leverage observations in the model from (e)?

### Note(s)

The data is a simulated data set containing sales of child car seats at 400 different stores. Information on the data set can be found on the [ISLP documentation](#).

Variable	Description
Sales	Unit sales (in thousands) at each location
CompPrice	Price charged by competitor at each location
Income	Community income level (in thousands of dollars)
Advertising	Local advertising budget for company at each location (in thousands of dollars)
Population	Population size in region (in thousands)
Price	Price company charges for car seats at each site
ShelveLoc	Factor with levels Bad, Good, Medium — quality of shelving location
Age	Average age of the local population
Education	Education level at each location
Urban	Factor (No/Yes) — whether store is in urban or rural location
US	Factor (No/Yes) — whether store is in the US or not

Table 1: Carseats Dataset Variables

---

## Solution

### ISLP 3.10 (a)

Table 2 shows the results of a linear model to predict Sales using Price, Urban, and US. The table was generated using the `statsmodels` library in Python.

<b>Dep. Variable:</b>	Sales	<b>R-squared:</b>	0.239
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.234
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	41.52
<b>Date:</b>	Tue, 13 Jan 2026	<b>Prob (F-statistic):</b>	2.39e-23
<b>Time:</b>	18:41:11	<b>Log-Likelihood:</b>	-927.66
<b>No. Observations:</b>	400	<b>AIC:</b>	1863.
<b>Df Residuals:</b>	396	<b>BIC:</b>	1879.
<b>Df Model:</b>	3		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>Intercept</b>	13.0435	0.651	20.036	0.000	11.764	14.323
<b>Price</b>	-0.0545	0.005	-10.389	0.000	-0.065	-0.044
<b>Urban_Yes</b>	-0.0219	0.272	-0.081	0.936	-0.556	0.512
<b>US_Yes</b>	1.2006	0.259	4.635	0.000	0.691	1.710

<b>Omnibus:</b>	0.676	<b>Durbin-Watson:</b>	1.912
<b>Prob(Omnibus):</b>	0.713	<b>Jarque-Bera (JB):</b>	0.758
<b>Skew:</b>	0.093	<b>Prob(JB):</b>	0.684
<b>Kurtosis:</b>	2.897	<b>Cond. No.</b>	628.

Table 2: OLS Regression Results

### ISLP 3.10 (b)

- **Price:** Sales is negatively related to price, with sales decreasing by 54 units per dollar increase in price
- **Urban\_Yes:** There is not a significant statistical relationship between sales and whether or not a store is in an urban location
- **US\_Yes:** There is a positive relationship between price and whether a store is located in the US or not. On average, a store located in the US will sell 1200 more units than if they were located outside the US.

### ISLP 3.10 (c)

$$\text{Sales} = -0.0545 * \text{Price} + -0.0219 * \text{Urban\_Yes} + 1.2006 * \text{US\_Yes}$$

### ISLP 3.10 (d)

The **Urban\_Yes** predictor has a P value of 0.936, meaning we can reject the null hypothesis for it. We can not reject the null hypothesis for any of the other variables.

**ISLP 3.10 (e)**

Table 3 shows the results of a linear model to predict Sales using Price and US\_Yes.

<b>Dep. Variable:</b>	Sales	<b>R-squared:</b>	0.239
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.235
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	62.43
<b>Date:</b>	Tue, 13 Jan 2026	<b>Prob (F-statistic):</b>	2.66e-24
<b>Time:</b>	20:11:19	<b>Log-Likelihood:</b>	-927.66
<b>No. Observations:</b>	400	<b>AIC:</b>	1861.
<b>Df Residuals:</b>	397	<b>BIC:</b>	1873.
<b>Df Model:</b>	2		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P >  t	[0.025	0.975]
<b>Intercept</b>	13.0308	0.631	20.652	0.000	11.790	14.271
<b>Price</b>	-0.0545	0.005	-10.416	0.000	-0.065	-0.044
<b>US_Yes</b>	1.1996	0.258	4.641	0.000	0.692	1.708

<b>Omnibus:</b>	0.666	<b>Durbin-Watson:</b>	1.912
<b>Prob(Omnibus):</b>	0.717	<b>Jarque-Bera (JB):</b>	0.749
<b>Skew:</b>	0.092	<b>Prob(JB):</b>	0.688
<b>Kurtosis:</b>	2.895	<b>Cond. No.</b>	607.

Table 3: OLS Regression Results for Model with Price and US\_Yes

**ISLP 3.10 (f)**

Models a and e fit the data equally poorly. Both have an  $R^2$  value of 0.239

**ISLP 3.10 (g)**

Table 4 shows the 95% confidence intervals for the coefficients in the model from (e).

	0	1
Intercept	11.790320	14.271265
Price	-0.064760	-0.044195
US_Yes	0.691520	1.707766

Table 4: 95% Confidence Intervals for Coefficients in Model from (e)

**ISLP 3.10 (h)**

Table 5 shows the high leverage points in the model from (e). The Cook's Distances are plotted in Figure 1.

index	high_leverage_points
42	0.043338
125	0.025966
155	0.016106
156	0.015356
159	0.015707
165	0.028567
171	0.021014
174	0.029687
191	0.018039
203	0.015356
208	0.018235
269	0.019195
272	0.018687
313	0.023165
315	0.017049
356	0.018279
365	0.017399
367	0.023707
383	0.016514
386	0.016555

Table 5: High Leverage Points in Model from (e)

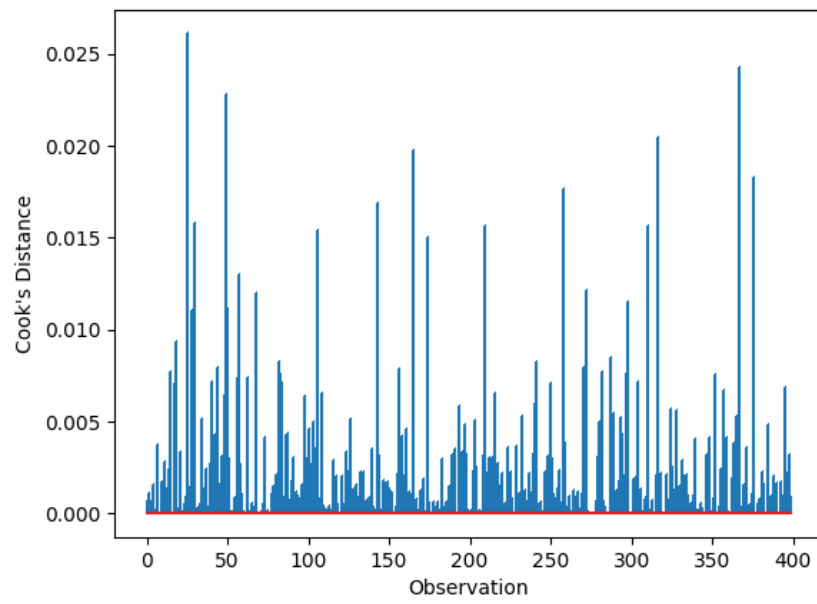


Figure 1: Cook's Distances for Model from (e)

## ISLP 3.14

This problem focuses on the *collinearity* problem.

(a) Perform the following commands in Python:

```
rng = np.random.default_rng(10)
x1 = rng.uniform(0, 1, size=100)
x2 = 0.5 * x1 + rng.normal(size=100) / 10
y = 2 + 2 * x1 + 0.3 * x2 + rng.normal(size=100)
```

The last line corresponds to creating a linear model in which  $y$  is a function of  $x_1$  and  $x_2$ . Write out the form of the linear model. What are the regression coefficients?

(b) What is the correlation between  $x_1$  and  $x_2$ ? Create a scatterplot displaying the relationship between the variables.

(c) Using this data, fit a least squares regression to predict  $y$  using  $x_1$  and  $x_2$ . Describe the results obtained. What are  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ ? How do these relate to the true  $\beta_0$ ,  $\beta_1$  and  $\beta_2$ ? Can you reject the null hypothesis  $H_0 : \beta_1 = 0$ ? How about the null hypothesis  $H_0 : \beta_2 = 0$ ?

(d) Now fit a least squares regression to predict  $y$  using only  $x_1$ . Comment on your results. Can you reject the null hypothesis  $H_0 : \beta_1 = 0$ ?

(e) Now fit a least squares regression to predict  $y$  using only  $x_2$ . Comment on your results. Can you reject the null hypothesis  $H_0 : \beta_2 = 0$ ?

(f) Do the results obtained in (c)-(e) contradict each other? Explain your answer.

(g) Suppose we obtain one additional observation, which was unfortunately mismeasured. We use the function `np.concatenate()` to add this additional observation to each of  $x_1$ ,  $x_2$  and  $y$ .

```
x1 = np.concatenate([x1, [0.1]])
x2 = np.concatenate([x2, [0.8]])
y = np.concatenate([y, [6]])
```

Re-fit the linear models from (c) to (e) using this new data. What effect does this new observation have on the each of the models? In each model, is this observation an outlier? A high-leverage point? Both? Explain your answers.

---

## Solution

### ISLP 3.14 (a)

The linear model has the form

$$y = 2 + 2x_1 + 0.3x_2 + \varepsilon$$

$$\beta_0 = 2$$

$$\beta_1 = 2$$

$$\beta_2 = 0.3$$

**ISLP 3.14 (b)**

$$\text{corr}(x_1, x_2) = \frac{\text{Cov}(x_1, x_2)}{\sigma_{x_1} * \sigma_{x_2}} = \frac{E[(x_1 - \bar{x}_1)(x_2 - \bar{x}_2)]}{\sigma_{x_1} * \sigma_{x_2}} = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \cdot \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}}$$

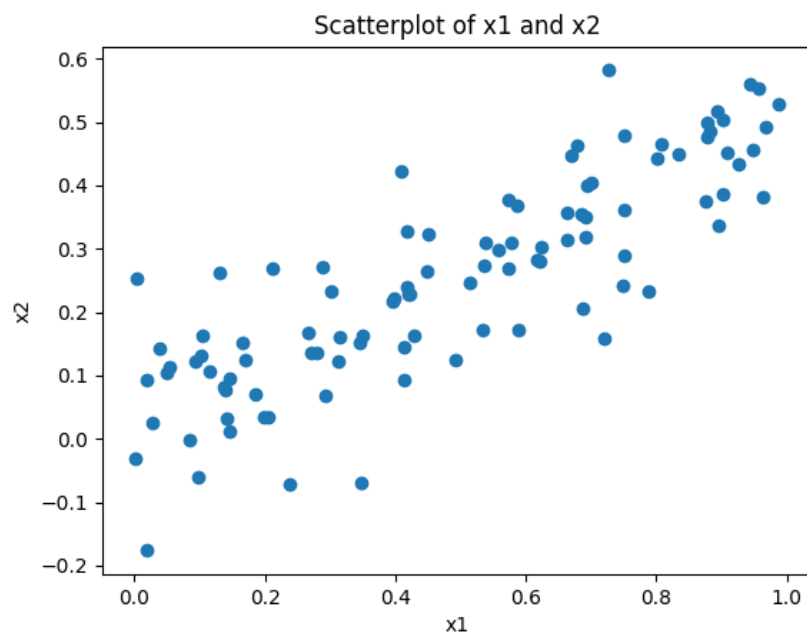


Figure 2: Scatterplot of x1 and x2

**ISLP 3.14 (c)**



<b>Dep. Variable:</b>	y	<b>R-squared:</b>	0.291
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.276
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	19.89
<b>Date:</b>	Tue, 13 Jan 2026	<b>Prob (F-statistic):</b>	5.76e-08
<b>Time:</b>	21:24:01	<b>Log-Likelihood:</b>	-130.62
<b>No. Observations:</b>	100	<b>AIC:</b>	267.2
<b>Df Residuals:</b>	97	<b>BIC:</b>	275.1
<b>Df Model:</b>	2		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P>  t	[0.025	0.975]
<b>const</b>	1.9579	0.190	10.319	0.000	1.581	2.334
<b>x1</b>	1.6154	0.527	3.065	0.003	0.569	2.661
<b>x2</b>	0.9428	0.831	1.134	0.259	-0.707	2.592

<b>Omnibus:</b>	0.051	<b>Durbin-Watson:</b>	1.964
<b>Prob(Omnibus):</b>	0.975	<b>Jarque-Bera (JB):</b>	0.041
<b>Skew:</b>	-0.036	<b>Prob(JB):</b>	0.979
<b>Kurtosis:</b>	2.931	<b>Cond. No.</b>	11.9

Table 6: OLS Regression Results for Model with x1 and x2

The linear model obtained from the OLS is

$$y = 1.9579 + 1.6154x_1 + 0.9428x_2$$

and

Variable	True	OLS Model
$\beta_0$	2.0	1.9579
$\beta_1$	2.0	1.6154
$\beta_2$	0.3	0.9428

Table 7: Coefficients for Model with x1 and x2

We can reject the null hypothesis  $H_0 : \beta_1 = 0$  as its t value is large ( $t_{\beta_1} = 3.065$ ) and P value is small ( $P_{\beta_1} = 0.003$ ). However, we can not reject the null hypothesis  $H_0 : \beta_2 = 0$  because its t value is less than 2 ( $t_{\beta_2} = 1.134$ ) along with a large P value ( $P_{\beta_2} = 0.259$ )

#### ISLP 3.14 (d)

Table 8 shows the results of the OLS regression to predict y using only x1. The goodness of fit of the least squares regression did not change meaningfully when we elmiited x2 as a predictor variable ( $R^2_{x1,x2} = 0.291$ ,  $R^2_{x1} = 0.281$ ). We can reject the null hypothesis  $H_0 : \beta_1 = 0$  due to the large t value and small P value for x1

	coef	std err	t	P>  t	[0.025	0.975]
<b>const</b>	1.9371	0.189	10.242	0.000	1.562	2.312
<b>x1</b>	2.0771	0.335	6.196	0.000	1.412	2.742
<b>R-squared:</b>	0.281		<b>Adj. R-squared:</b>	0.274		
<b>F-statistic:</b>	38.39		<b>Prob (F-statistic):</b>	1.37e-08		

Table 8: OLS Regression Results for Model with x1

#### ISLP 3.14 (e)

Table 9 shows the results of the OLS regression to predict y using only x2. The goodness of fit of the least squares regression with only x2 included did not change meaningfully when we elmiited x1 as a predictor variable ( $R^2_{x1,x2} = 0.291$ ,  $R^2_{x2} = 0.222$ ). We can reject the null hypothesis  $H_0 : \beta_2 = 0$  due to the large t value and small P value for x2

	coef	std err	t	P>  t	[0.025	0.975]
<b>const</b>	2.3239	0.154	15.124	0.000	2.019	2.629
<b>x2</b>	2.9103	0.550	5.291	0.000	1.819	4.002
<b>R-squared:</b>	0.222		<b>Adj. R-squared:</b>	0.214		
<b>F-statistic:</b>	27.99		<b>Prob (F-statistic):</b>	7.43e-07		

Table 9: OLS Regression Results for Model with x2

#### ISLP 3.14 (f)

They do not contradict each other, x1 and x2 explain the variance in y equally well (poorly) with  $R^2_{x1} = 0.281$  and  $R^2_{x2} = 0.222$ . Additionally the null hypothesis can not be rejected when only including 1 predictor, either x1 or x2. These outcomes can be explained by x1 and x2 being colinear (they are highly coorelated).

#### ISLP 3.14 (g)

Table 10 shows the results of the OLS regression to predict y using x1, x2, x1 only, and x2 only. Figure 3 shows some diagnostic plots for the full model.

Full Model				
	coef	t value	p value	std err
intercept	2.0618	10.7201	0.0000	0.1923
$x_1$	0.8575	1.8383	0.0690	0.4665
$x_2$	2.2663	3.2160	0.0018	0.7047
R-squared	0.2916	F-statistic		20.1727

$x_1$ Only Model				
	coef	t value	p value	std err
intercept	2.0739	10.3098	0.0000	0.2012
$x_1$	1.8760	5.2360	0.0000	0.3583
R-squared	0.2169	F-statistic		27.4159

$x_2$ Only Model				
	coef	t value	p value	std err
intercept	2.2840	15.0879	0.0000	0.1514
$x_2$	3.1458	6.0082	0.0000	0.5236
R-squared	0.2672	F-statistic		36.0984

Table 10: OLS Regression Results for Models with  $x_1$ ,  $x_2$ ,  $x_1$  only, and  $x_2$  only

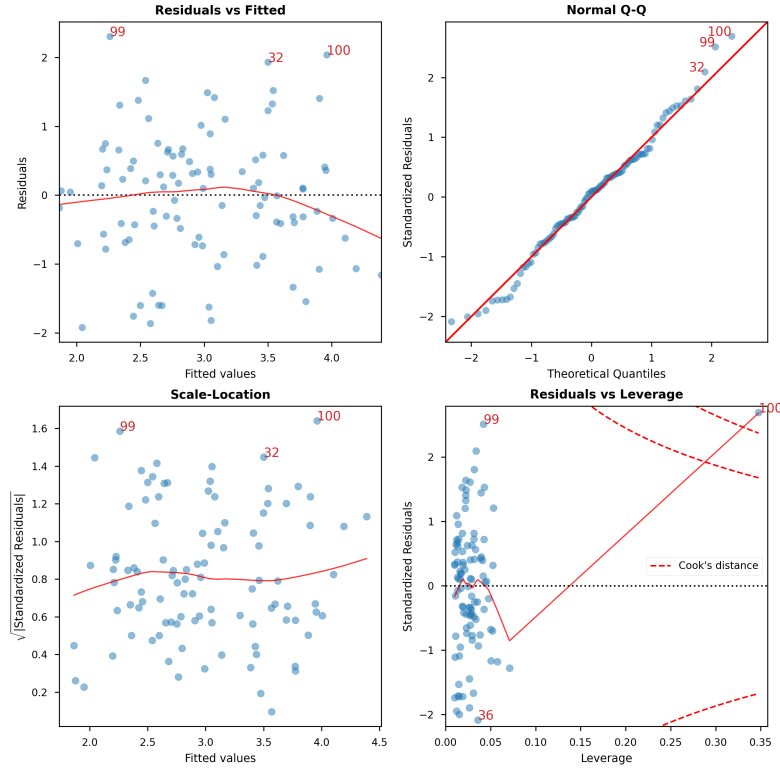


Figure 3: Diagnostic Plots for Full Model

With the added point, we can no longer reject the null hypothesis for  $\beta_1$ , as the t value is now 1.8383 and the P value is now 0.0690. We can now, however, reject the null hypothesis for  $\beta_2$ , as the t value is now 3.2160 and the P value is now 0.0018. The added point is a high-leverage point and an outlier in the full model. All models fit the data equally well (poorly), with  $R^2$  values all less than 0.3.

The added point does have a high leverage (0.347) but is not an outlier because its Cook's Distance is not greater than 1 (Cook's Distance = 0.316).

### ISLP 3.15 (a, b, d)

This problem involves the Boston data set, which we saw in the lab for this chapter. We will now try to predict per capita crime rate using the other variables in this data set. In other words, per capita crime rate is the response, and the other variables are the predictors.

(a) For each predictor, fit a simple linear regression model to predict the response. Describe your results. In which of the models is there a statistically significant association between the predictor and the response? Create some plots to back up your assertions.

(b) Fit a multiple regression model to predict the response using all of the predictors. Describe your results. For which predictors can we reject the null hypothesis  $H_0 : \beta_j = 0$ ?

(d) Is there evidence of non-linear association between any of the predictors and the response? To answer this question, for each predictor  $X$ , fit a model of the form

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon \quad (1)$$

---

## Solution

### ISLP 3.15 (a)

Figure 4 shows the scatterplots for each predictor. Figure 5 shows the residuals for each predictor. Table 11 shows the R-squared for each predictor. All R-squared values are less than 0.4, indicating that the univariate models are not very good at predicting the response.

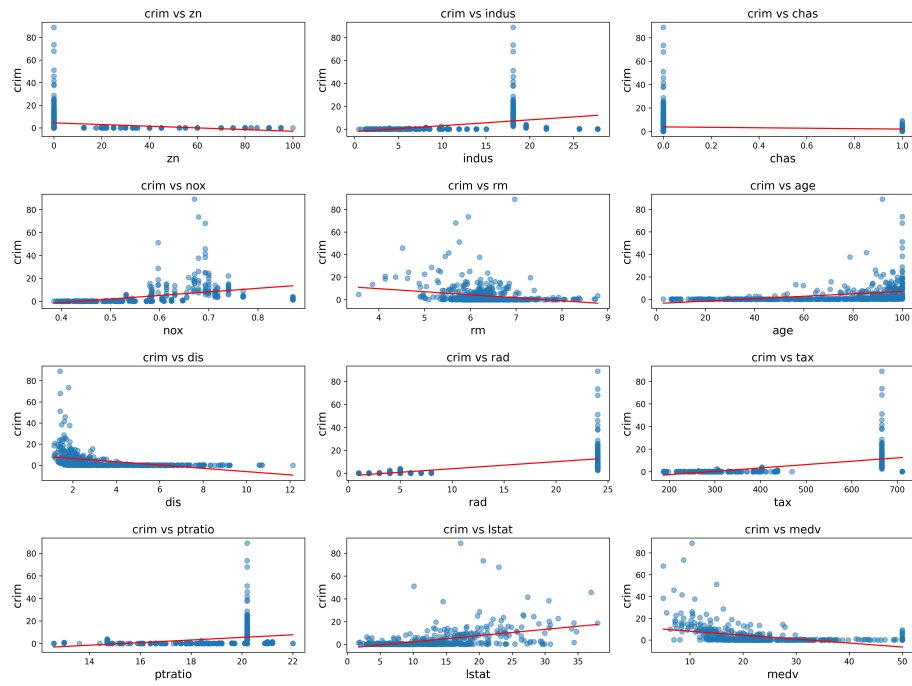


Figure 4: Scatterplots for Each Predictor

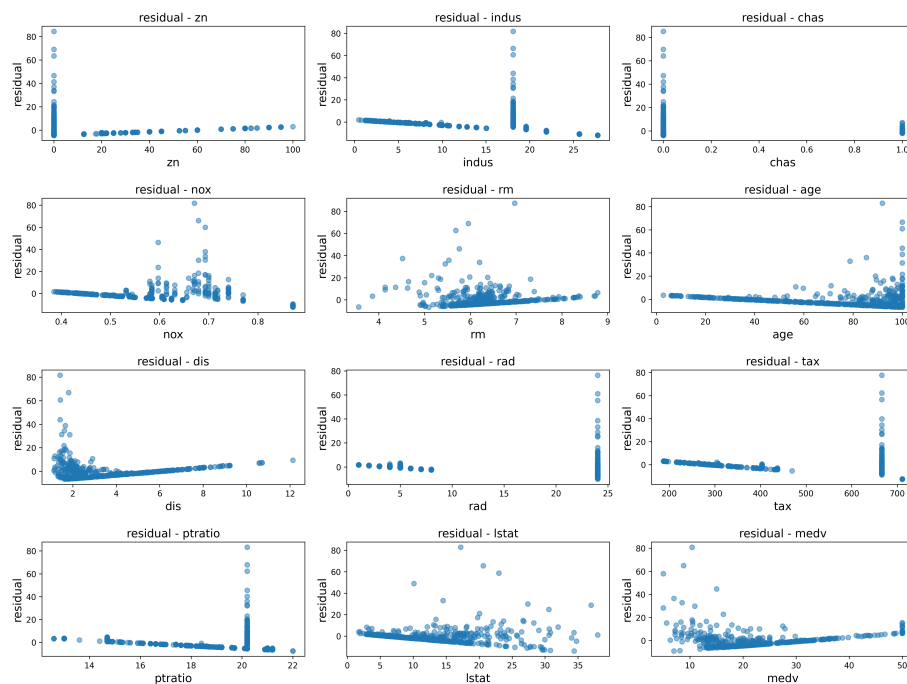


Figure 5: Residuals for Each Predictor

predictor	r_squared
rad	0.391257
tax	0.339614
lstat	0.207591
nox	0.177217
indus	0.165310
medv	0.150780
dis	0.144149
age	0.124421
ptratio	0.084068
rm	0.048069
zn	0.040188
chas	0.003124

Table 11: R-squared for Each Predictor

### ISLP 3.15 (b)

Table 12 shows the results of the OLS regression to predict crim using all of the predictors. We can reject the null hypothesis, with a 95% confidence interval  $\alpha = 0.05$ , for zn, dis, rad, and medv.

We can not reject the null hypothesis for all other predictors.

<b>Dep. Variable:</b>	crim	<b>R-squared:</b>	0.449
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.436
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	33.52
<b>Date:</b>	Tue, 13 Jan 2026	<b>Prob (F-statistic):</b>	2.03e-56
<b>Time:</b>	23:53:03	<b>Log-Likelihood:</b>	-1655.4
<b>No. Observations:</b>	506	<b>AIC:</b>	3337.
<b>Df Residuals:</b>	493	<b>BIC:</b>	3392.
<b>Df Model:</b>	12		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P>  t	[0.025	0.975]
<b>Intercept</b>	13.7784	7.082	1.946	0.052	-0.136	27.693
<b>zn</b>	0.0457	0.019	2.433	0.015	0.009	0.083
<b>indus</b>	-0.0584	0.084	-0.698	0.486	-0.223	0.106
<b>chas</b>	-0.8254	1.183	-0.697	0.486	-3.150	1.500
<b>nox</b>	-9.9576	5.290	-1.882	0.060	-20.351	0.436
<b>rm</b>	0.6289	0.607	1.036	0.301	-0.564	1.822
<b>age</b>	-0.0008	0.018	-0.047	0.962	-0.036	0.034
<b>dis</b>	-1.0122	0.282	-3.584	0.000	-1.567	-0.457
<b>rad</b>	0.6125	0.088	6.997	0.000	0.440	0.784
<b>tax</b>	-0.0038	0.005	-0.730	0.466	-0.014	0.006
<b>ptratio</b>	-0.3041	0.186	-1.632	0.103	-0.670	0.062
<b>lstat</b>	0.1388	0.076	1.833	0.067	-0.010	0.288
<b>medv</b>	-0.2201	0.060	-3.678	0.000	-0.338	-0.103

<b>Omnibus:</b>	663.436	<b>Durbin-Watson:</b>	1.516
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	80856.852
<b>Skew:</b>	6.579	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	63.514	<b>Cond. No.</b>	1.24e+04

Table 12: OLS Regression Results for Multiple Regression Model

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 1.24e+04. This might indicate that there are strong multicollinearity or other numerical problems.

### ISLP 3.15 (d)

Tables 13 - 15 show the results of the polynomial regressions for each predictor. Figure 6 and figure 7 show the scatter and residuals plots for the polynomial regression models respectively.

Looking at the scatter plots, we can see that none of the variables have great non-linear fits to the response. Looking exclusively at the P values for the different polynomial models, we can not reject a non-linear relationship for indius, nox, age, dis, ptratio, and medv.



Model for zn:				
	coef	t value	p value	std err
intercept	4.8461	11.1922	0.0000	0.4330
zn	-0.3322	-3.0252	0.0026	0.1098
zn <sup>2</sup>	0.0065	1.6791	0.0938	0.0039
zn <sup>3</sup>	-0.0000	-1.2030	0.2295	0.0000
R-squared	0.0582	F-statistic		10.3485

Model for indus:				
	coef	t value	p value	std err
intercept	3.6626	2.3269	0.0204	1.5740
indus	-1.9652	-4.0773	0.0001	0.4820
indus <sup>2</sup>	0.2519	6.4070	0.0000	0.0393
indus <sup>3</sup>	-0.0070	-7.2920	0.0000	0.0010
R-squared	0.2597	F-statistic		58.6883

Model for chas:				
	coef	t value	p value	std err
intercept	3.7444	9.4530	0.0000	0.3961
chas	-0.6309	-1.2567	0.2094	0.5020
chas <sup>2</sup>	-0.6309	-1.2567	0.2094	0.5020
chas <sup>3</sup>	-0.6309	-1.2567	0.2094	0.5020
R-squared	0.0031	F-statistic		1.5794

Model for nox:				
	coef	t value	p value	std err
intercept	233.0866	6.9282	0.0000	33.6431
nox	-1279.3713	-7.5082	0.0000	170.3975
nox <sup>2</sup>	2248.5441	8.0334	0.0000	279.8993
nox <sup>3</sup>	-1245.7029	-8.3446	0.0000	149.2816
R-squared	0.2970	F-statistic		70.6867

Table 13: Polynomial Regression Results for Each Predictor

Model for rm:				
	coef	t value	p value	std err
intercept	112.6246	1.7457	0.0815	64.5172
rm	-39.1501	-1.2503	0.2118	31.3115
rm <sup>2</sup>	4.5509	0.9084	0.3641	5.0099
rm <sup>3</sup>	-0.1745	-0.6615	0.5086	0.2637
R-squared	0.0678	F-statistic		12.1677

Model for age:				
	coef	t value	p value	std err
intercept	-2.5488	-0.9204	0.3578	2.7691
age	0.2737	1.4683	0.1427	0.1864
age <sup>2</sup>	-0.0072	-1.9878	0.0474	0.0036
age <sup>3</sup>	0.0001	2.7237	0.0067	0.0000
R-squared	0.1742	F-statistic		35.3061

Model for dis:				
	coef	t value	p value	std err
intercept	30.0476	12.2850	0.0000	2.4459
dis	-15.5544	-8.9600	0.0000	1.7360
dis <sup>2</sup>	2.4521	7.0783	0.0000	0.3464
dis <sup>3</sup>	-0.1186	-5.8135	0.0000	0.0204
R-squared	0.2778	F-statistic		64.3741

Model for rad:				
	coef	t value	p value	std err
intercept	-0.6055	-0.2954	0.7678	2.0501
rad	0.5127	0.4913	0.6234	1.0436
rad <sup>2</sup>	-0.0752	-0.5061	0.6130	0.1485
rad <sup>3</sup>	0.0032	0.7031	0.4823	0.0046
R-squared	0.4000	F-statistic		111.5727

Table 14: Polynomial Regression Results for Each Predictor

Model for tax:				
	coef	t value	p value	std err
intercept	19.1836	1.6263	0.1045	11.7955
tax	-0.1533	-1.6023	0.1097	0.0957
tax <sup>2</sup>	0.0004	1.4877	0.1375	0.0002
tax <sup>3</sup>	-0.0000	-1.1668	0.2439	0.0000
R-squared	0.3689	F-statistic		97.8047

Model for ptratio:				
	coef	t value	p value	std err
intercept	477.1840	3.0434	0.0025	156.7950
ptratio	-82.3605	-2.9793	0.0030	27.6439
ptratio <sup>2</sup>	4.6353	2.8821	0.0041	1.6083
ptratio <sup>3</sup>	-0.0848	-2.7433	0.0063	0.0309
R-squared	0.1138	F-statistic		21.4839

Model for lstat:				
	coef	t value	p value	std err
intercept	1.2010	0.5920	0.5541	2.0286
lstat	-0.4491	-0.9660	0.3345	0.4649
lstat <sup>2</sup>	0.0558	1.8522	0.0646	0.0301
lstat <sup>3</sup>	-0.0009	-1.5170	0.1299	0.0006
R-squared	0.2179	F-statistic		46.6294

Model for medv:				
	coef	t value	p value	std err
intercept	53.1655	15.8405	0.0000	3.3563
medv	-5.0948	-11.7438	0.0000	0.4338
medv <sup>2</sup>	0.1555	9.0455	0.0000	0.0172
medv <sup>3</sup>	-0.0015	-7.3120	0.0000	0.0002
R-squared	0.4202	F-statistic		121.2721

Table 15: Polynomial Regression Results for Each Predictor

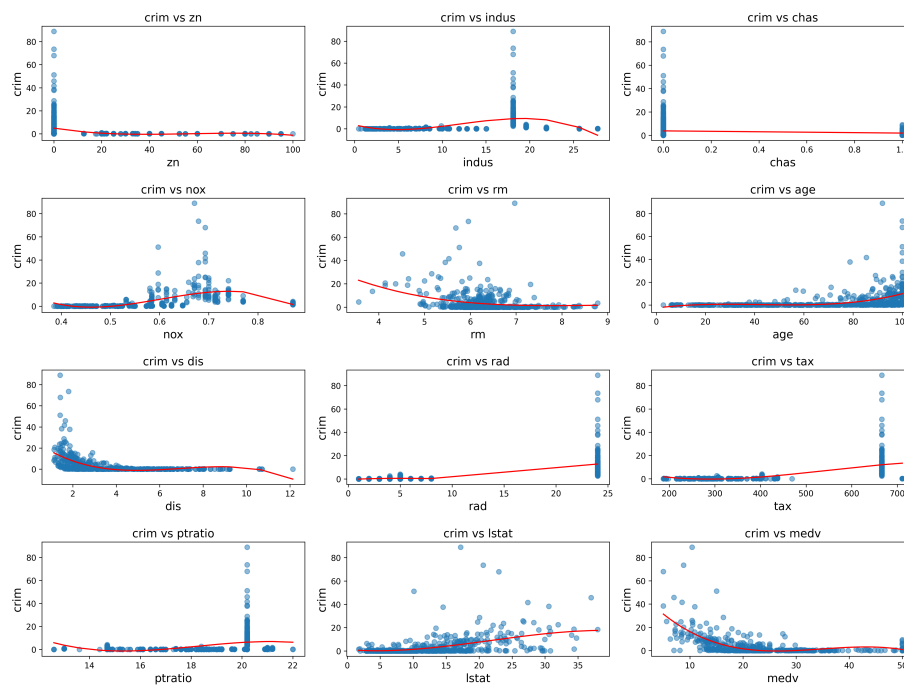


Figure 6: Scatterplots for Polynomial Regression Models

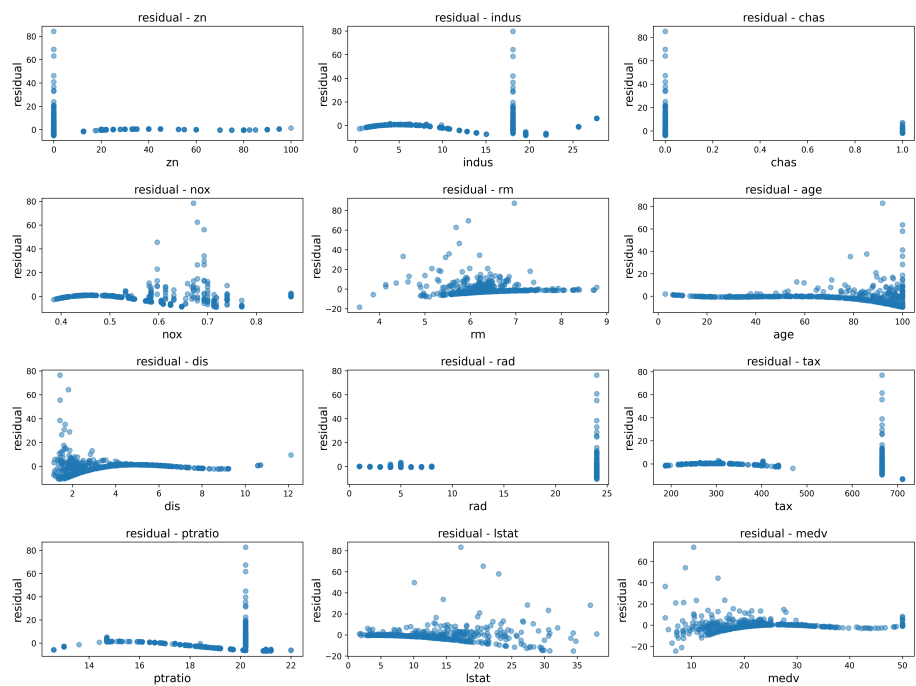


Figure 7: Residuals for Polynomial Regression Models

## ESL 3.17

Repeat the analysis of Table 3.3 on the spam data discussed in Chapter 1. Include LS, Best Subset, Ridge regression, and Lasso. (skip PCR and PLS)

**Note(s)**

- **Best Subset:** Best-subset selection drops all variables with coefficients smaller than the Mth largest; this is a form of “hard-thresholding.”
- **Ridge regression:**

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad \text{with } \lambda \geq 1$$

which can be written in matrix form as

$$\text{RSS}(\lambda) = (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) + \lambda \beta^T \beta$$

with

$$\hat{\beta}^{\text{ridge}} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$

- **Lasso:** Similar to ridge but with an  $L_1$  penalty instead of  $L_2$ :

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \right\}$$

subject to  $\sum_{j=1}^p |\beta_j| \leq t$

or equivalently in \*Lagrangian form\*

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^N \left( y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad \text{with } \lambda \geq 1$$

**NOTE:** t should be adaptively chosen to minimize an estimate of expected prediction error.

**NOTE:** Ridge regression does a proportional shrinkage. Lasso translates each coefficient by a constant factor  $\lambda$ , truncating at zero (“soft thresholding”)

---

## Solution

Figure 8 shows the best subset score for each model. Figure 9 shows the coefficients for each model. Table 16 shows the coefficients for each model, mirroring table 3.3 in the ESL textbook.

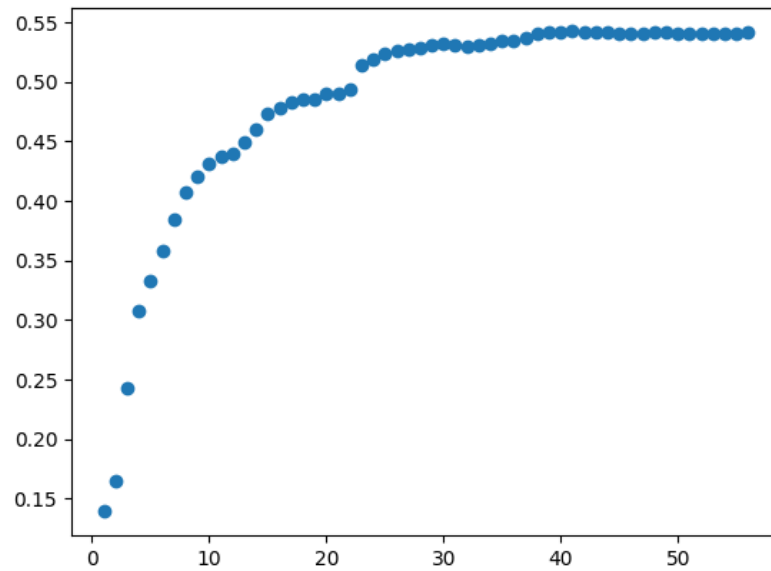


Figure 8: Best Subset Score

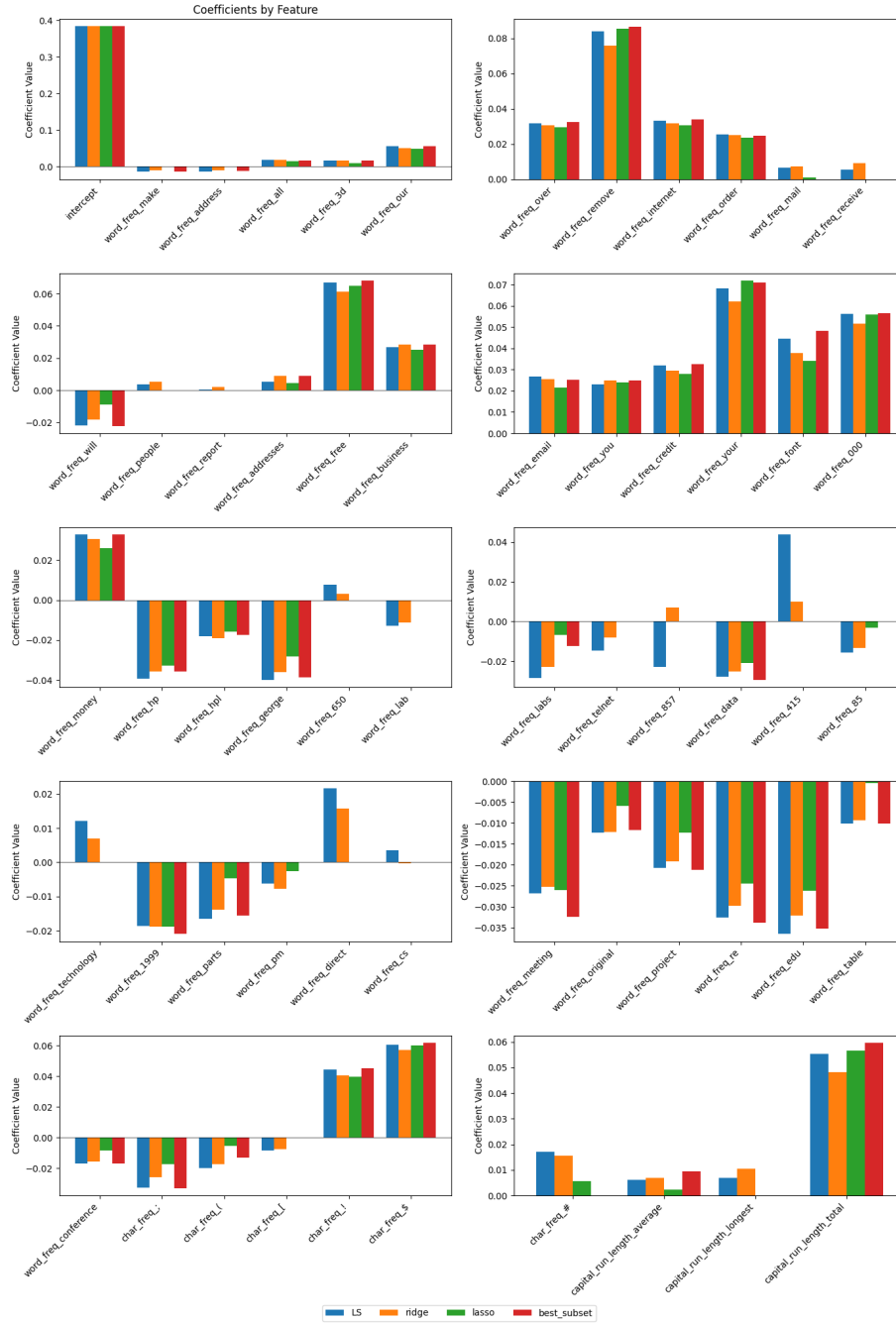


Figure 9: Coefficients



	LS	ridge	lasso	best_subset
intercept	0.3839	0.3839	0.3839	0.3839
word_freq_make	-0.0149	-0.0100	0.0000	-0.0137
word_freq_address	-0.0133	-0.0111	0.0000	-0.0123
word_freq_all	0.0175	0.0180	0.0139	0.0169
word_freq_3d	0.0170	0.0157	0.0089	0.0169
word_freq_our	0.0551	0.0498	0.0480	0.0551
word_freq_over	0.0318	0.0306	0.0297	0.0324
word_freq_remove	0.0839	0.0760	0.0857	0.0866
word_freq_internet	0.0333	0.0316	0.0305	0.0340
word_freq_order	0.0254	0.0252	0.0236	0.0249
word_freq_mail	0.0065	0.0073	0.0011	0.0000
word_freq_receive	0.0053	0.0090	0.0000	0.0000
word_freq_will	-0.0219	-0.0183	-0.0088	-0.0222
word_freq_people	0.0038	0.0055	0.0000	0.0000
word_freq_report	0.0005	0.0019	0.0000	0.0000
word_freq_addresses	0.0055	0.0089	0.0044	0.0089
word_freq_free	0.0669	0.0613	0.0648	0.0682
word_freq_business	0.0269	0.0282	0.0250	0.0283
word_freq_email	0.0268	0.0254	0.0216	0.0252
word_freq_you	0.0231	0.0248	0.0238	0.0249
word_freq_credit	0.0318	0.0295	0.0280	0.0326
word_freq_your	0.0681	0.0620	0.0718	0.0709
word_freq_font	0.0445	0.0378	0.0340	0.0483
word_freq_000	0.0562	0.0516	0.0559	0.0564
word_freq_money	0.0328	0.0305	0.0259	0.0327
word_freq_hp	-0.0394	-0.0356	-0.0329	-0.0357
word_freq_hpl	-0.0179	-0.0189	-0.0157	-0.0176
word_freq_george	-0.0400	-0.0359	-0.0283	-0.0387
word_freq_650	0.0078	0.0033	0.0000	0.0000
word_freq_lab	-0.0129	-0.0113	0.0000	0.0000
word_freq_labs	-0.0284	-0.0227	-0.0069	-0.0122
word_freq_telnet	-0.0145	-0.0082	0.0000	0.0000
word_freq_857	-0.0227	0.0071	0.0000	0.0000
word_freq_data	-0.0278	-0.0251	-0.0207	-0.0295
word_freq_415	0.0438	0.0099	0.0000	0.0000
word_freq_85	-0.0156	-0.0132	-0.0030	0.0000
word_freq_technology	0.0121	0.0069	0.0000	0.0000
word_freq_1999	-0.0186	-0.0189	-0.0188	-0.0210
word_freq_parts	-0.0165	-0.0138	-0.0048	-0.0155
word_freq_pm	-0.0062	-0.0077	-0.0027	0.0000
word_freq_direct	0.0216	0.0157	0.0000	0.0000
word_freq_cs	0.0035	-0.0003	0.0000	0.0000
word_freq_meeting	-0.0268	-0.0253	-0.0260	-0.0323
word_freq_original	-0.0123	-0.0121	-0.0059	-0.0116

	LS	ridge	lasso	best_subset
word_freq_project	-0.0207	-0.0192	-0.0123	-0.0212
word_freq_re	-0.0326	-0.0297	-0.0245	-0.0338
word_freq_edu	-0.0365	-0.0321	-0.0262	-0.0353
word_freq_table	-0.0102	-0.0093	-0.0004	-0.0100
word_freq_conference	-0.0171	-0.0157	-0.0084	-0.0169
char_freq_;	-0.0328	-0.0259	-0.0176	-0.0331
char_freq_(	-0.0199	-0.0174	-0.0053	-0.0131
char_freq_[	-0.0085	-0.0074	0.0000	0.0000
char_freq_!	0.0445	0.0406	0.0398	0.0452
char_freq_\$	0.0607	0.0571	0.0600	0.0619
char_freq_#	0.0171	0.0154	0.0055	0.0000
capital_run_length_avg	0.0062	0.0069	0.0023	0.0094
capital_run_length_long	0.0069	0.0104	0.0000	0.0000
capital_run_length_total	0.0554	0.0483	0.0565	0.0596
test R2	0.5419	0.5450	0.5302	0.5420
test error	0.1114	0.1107	0.1143	0.1114

Table 16: ESL 3.17: Model Coefficients