Stat 206: Linear Models

Lecture 11

Nov. 4, 2015

Polynomial Regression

Polynomial regression models are among the most commonly used models to describe a regression relation.

- Polynomial regression models are very flexible and are easy to fit.
- Polynomial models with higher than third-order terms are rarely employed in practice.
  - They often lead to estimators.
  - They may fit the observed data very well, but generalize well to new observations, a phenomena called .

Polynomial Regression

Polynomial regression models are among the most commonly used models to describe a curvilinear regression relation.

- Polynomial regression models are very flexible and are easy to fit.
- Polynomial models with higher than third-order terms are rarely employed in practice.
    - They often lead to highly variable estimators.
    - They may fit the observed data very well, but may not generalize well to new observations, a phenomena called *overfitting*.

Second-Order Model with One Predictor

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1(X_i - \overline{X}) + \beta_2(X_i - \overline{X})^2 + \epsilon_i \\
&= \beta_0 + \beta_1\tilde{X}_i + \beta_2\tilde{X}_i^2 + \epsilon_i, \quad i = 1, \cdots, n,
\end{aligned}
$$

where $\tilde{X}_i = X_i - \overline{X}$ is the centered value of the predictor variable in the ith case.

- Centering often between the linear term X and the quadratic term X^2 substantially *(Why?)* and thus improves numerical accuracy. *Will centering change the fitted regression function?*
- The response function is a parabola:
  - β_0 is the mean response when
    .
  - β_1 is called the and β_2 is called the
    .

Second-Order Model with One Predictor

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(X_i - \overline{X}) + \beta_2(X_i - \overline{X})^2 + \epsilon_i \\ &= \beta_0 + \beta_1 \tilde{X}_i + \beta_2 \tilde{X}_i^2 + \epsilon_i, \quad i = 1, \cdots, n, \end{aligned}$$

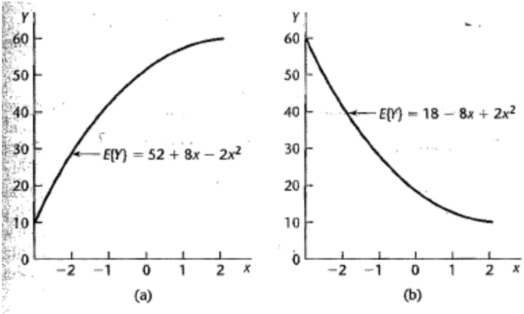
where $\tilde{X}_i = X_i - \overline{X}$ is the centered value of the predictor variable in the *i*th case.

- Centering often reduces the correlation between the linear term X and the quadratic term X^2 substantially *(Why?)* and thus improves numerical accuracy. *Will centering change the fitted regression function?*
- The response function is a parabola:

$$\begin{aligned} E(Y) &= \beta_0 + \beta_1(X - \overline{X}) + \beta_2(X - \overline{X})^2 \\ &= \beta_0 + \beta_1 \tilde{X} + \beta_2 \tilde{X}^2. \end{aligned}$$

  - β_0 is the mean response when $\tilde{X} = 0$, i.e. when $X = \overline{X}$.
  - β_1 is called the *linear effect coefficient* and β_2 is called the *quadratic effect coefficient*.

Figure: Examples of quadratic response functions.



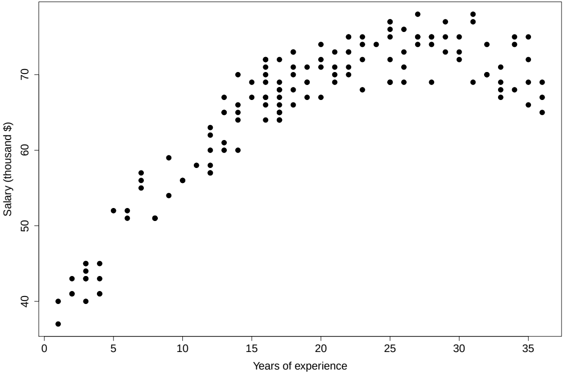
Salary

Professional organizations regularly survey their members for information concerning salaries, pensions, and conditions of employment. One goal is to relate salary to years of experience. This data has years of experience (X) and salary (Y) on 143 cases. [1]

Case	Salary($) Y	Experience(Years) X
1	71	26
2	69	19
3	73	22
...	...	...
141	67	16
142	71	20
143	69	31

[1] Source of data: Tryfos (1998): Methods for business analysis and forecasting

Figure: Scatter plot of salary versus years of experience



The regression relation between salary and years of experience appears to be curvilinear.

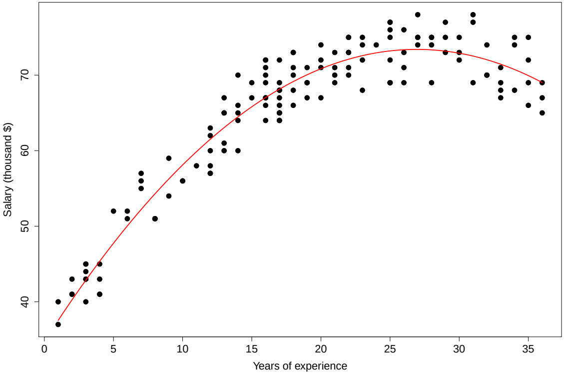
Case	Salary($)	Experience(Years)	Experience^2	Centered_Experience	(Centered_Experience)^2
Y	X	X^2	centered_X		(centered_X)^2
1	71	26	676	7.14	50.98
2	69	19	361	0.14	0.02
3	73	22	484	3.14	9.86
4	69	17	289	-1.86	3.46
5	65	13	169	-5.86	34.34
6	75	25	625	6.14	37.70
...	...	...	...	...	...

Average years of experience $\overline{X} = 18.86$. Correlation coefficient between X and X^2 is 0.965 and correlation coefficient between $\tilde{X}$ and $\tilde{X}^2$ is -0.0414.

Salary: Second-Order Model

```
> salary.c=salary
> salary.c[,"Experience"]=salary[,"Experience"]-mean(salary[,"Experience"]) ## center the X variable
> fitc=lm(Salary˜ Experience+I(Experienceˆ2), data=salary.c) ## fit a second-order model
> summary(fitc)

Call:
lm(formula = Salary ˜ Experience + I(Experienceˆ2), data = salary.c)

Residuals:
Min      1Q  Median      3Q     Max
-4.5786 -2.3573  0.0957  2.0171  5.5176

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)     69.927208   0.323090  216.43   <2e-16 ***
Experience       0.861177   0.024957   34.51   <2e-16 ***
I(Experienceˆ2) -0.053316   0.002477  -21.53   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.817 on 140 degrees of freedom
Multiple R-squared: 0.9247,     Adjusted R-squared: 0.9236
F-statistic: 859.3 on 2 and 140 DF,  p-value: < 2.2e-16
```

Figure: Fitted response function:
$y = 69.93 + 0.861 \times (X - 18.86) - 0.0533 \times (X - 18.86)^2$



Second-Order Model with Two Predictors

where $\tilde{X}_{i1} = X_{i1} - \overline{X}_1$, $\tilde{X}_{i2} = X_{i2} - \overline{X}_2$.

- Response function is a conic section:

$$E(Y) = \beta_0 + \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \beta_{11} \tilde{X}_1^2 + \beta_{22} \tilde{X}_2^2 + \beta_{12} \tilde{X}_1 \tilde{X}_2.$$

- This model contains separate and terms for each of the two predictors.
- It also contains a term representing the between the two predictors.
- β_{12} is called the .

Second-Order Model with Two Predictors

$$Y_i = \beta_0 + \beta_1 \tilde{X}_{i1} + \beta_2 \tilde{X}_{i2} + \beta_{11} \tilde{X}_{i1}^2 + \beta_{22} \tilde{X}_{i2}^2 + \beta_{12} \tilde{X}_{i1} \tilde{X}_{i2} + \epsilon_i, i = 1, \cdots, n,$$

where $\tilde{X}_{i1} = X_{i1} - \overline{X}_1$, $\tilde{X}_{i2} = X_{i2} - \overline{X}_2$.

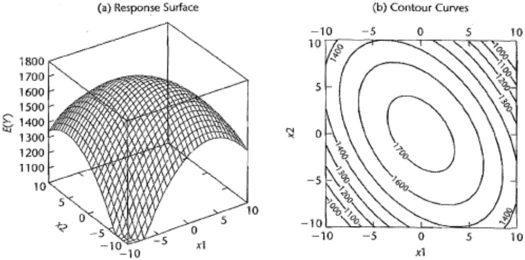
- Response function is a conic section:

$$E(Y) = \beta_0 + \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \beta_{11} \tilde{X}_1^2 + \beta_{22} \tilde{X}_2^2 + \beta_{12} \tilde{X}_1 \tilde{X}_2.$$

- This model contains separate linear and quadratic terms for each of the two predictors.
- It also contains a cross-product term representing the interaction between the two predictors.
- β_{12} is called the *interaction effect coefficient*.

Figure: A quadratic response surface.

FIGURE 8.3 Example of a Quadratic Response Surface—$E\{Y\} = 1{,}740 - 4x_1^2 - 3x_2^2 - 3x_1x_2$.



The contour curves show various combinations of the values of the two predictors that yield the same value of the response function.

Coefficient of Partial Determination

It measures the marginal contribution in proportional reduction in SSE by adding one X variable into a model.

- Definition.

$$
\begin{aligned}
& R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1} \\
:= \quad & \frac{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1}) - SSE(X_1,\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})} \\
= \quad & \frac{SSR(X_j|X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}
\end{aligned}
$$

- Coefficients of partial determination are in between

  .

- For example, $R^2_{Y,1|2} =$ is

Coefficient of Partial Determination

It measures the marginal contribution in proportional reduction in SSE by adding one X variable into a model.

- Definition.

$$
\begin{aligned}
& R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1} \\
:= \quad & \frac{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1}) - SSE(X_1,\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})} \\
= \quad & \frac{SSR(X_j|X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}{SSE(X_1,\cdots,X_{j-1},X_{j+1},\cdots,X_{p-1})}
\end{aligned}
$$

- Coefficients of partial determination are in between 0 and 1.
- For example, $R^2_{Y,1|2} = \frac{SSR(X_1|X_2)}{SSE(X_2)}$ is the proportional reduction in SSE by including X_1 into the model with X_2.

Body Fat

- From R outputs, we can obtain a number of coefficients of partial determination. E.g.:

$$R^2_{Y,2|1} =$$

$$R^2_{Y,1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{113.42 - 109.95}{113.43} = 3.1\%.$$

$$R^2_{Y,3|12} =$$

- When X_2 is added to the model containing X_1, SSE is reduced by ; When X_1 is added to the model containing X_2, SSE is reduced by ; When X_3 is added to the model containing X_1, X_2, SSE is reduced by .

R outputs

Body Fat

- From R outputs, we can obtain a number of coefficients of partial determination. E.g.:

$$R^2_{Y,2|1} = \frac{SSE(X_1) - SSE(X_1, X_2)}{SSE(X_1)} = \frac{143.12 - 109.95}{143.12} = 23.2\%.$$

$$R^2_{Y,1|2} = \frac{SSE(X_2) - SSE(X_1, X_2)}{SSE(X_2)} = \frac{113.42 - 109.95}{113.43} = 3.1\%.$$

$$R^2_{Y,3|12} = \frac{SSR(X_3|X_1, X_2)}{SSE(X_1, X_2)} = \frac{11.55}{109.95} = 10.5\%.$$

- When X_2 is added to the model containing X_1, SSE is reduced by 23.2%; When X_1 is added to the model containing X_2, SSE is reduced by 3.1%; When X_3 is added to the model containing X_1, X_2, SSE is reduced by 10.5%.

◂ R outputs

Interpretation of Coefficient of Partial Determination

- The ESS, $SSR(X_j|X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_{p-1})$, is the SSR when regressing the residuals $e(Y|X_{-(j)}) = Y - \hat{Y}(X_{-(j)})$ to the residuals $e(X_j|X_{-(j)}) = X_j - \hat{X}_j(X_{-(j)})$, where $X_{-(j)} = \{X_l : 1 \le l \ne j \le p\}$.
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ is the

  between the two sets of residuals obtained by regressing Y and X_j to the rest of variables $X_{-(j)}$, respectively.
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ measures the linear association between Y and X_j after

  have been adjusted for.

Interpretation of Coefficient of Partial Determination

- The ESS, $SSR(X_j|X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_{p-1})$, is the SSR when regressing the residuals $e(Y|X_{-(j)}) = Y - \hat{Y}(X_{-(j)})$ to the residuals $e(X_j|X_{-(j)}) = X_j - \hat{X}_j(X_{-(j)})$, where $X_{-(j)} = \{X_l : 1 \leq l \neq j \leq p\}$.
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ is the coefficient of simple determination (i.e., the squared correlation coefficient)between the two sets of residuals obtained by regressing Y and X_j to the rest of variables $X_{-(j)}$, respectively.
- So $R^2_{Y,j|1,\cdots,j-1,j+1,\cdots,p-1}$ measures the linear association between Y and X_j after the linear effects of $X_{-(j)}$ have been adjusted for.

Example. $R^2_{Y,1|2}$.

- Regress Y on X_2: $e_i(Y|X_2) = Y_i - \widehat{Y}_i(X_2), \;\; i = 1, \cdots n.$
- Regress X_1 on X_2: $e_i(X_1|X_2) = X_{i1} - \widehat{X}_{i1}(X_2), \;\; i = 1, \cdots n.$
- $R^2_{Y1|2}$ equals to the coefficient of simple determination between $e_i(Y|X_2)$ and $e_i(X_1|X_2)$.
- It measures the linear association between Y and X_1 after the linear effects of X_2 have been adjusted for.

Partial Correlations

The **signed** square-root of a coefficient of partial determination is called a partial correlation.

- The sign is the same as the sign of the corresponding fitted regression coefficient.
- Partial correlation is the
  between the .
- Partial correlations can be used to find the “best” X variable to be added next for inclusion in the regression model.

Partial Correlations

The **signed** square-root of a coefficient of partial determination is called a partial correlation.

- The sign is the same as the sign of the corresponding fitted regression coefficient.
- Partial correlation is the correlation coefficient between the two respective sets of residuals.
- Partial correlations can be used to find the “best” X variable to be added next for inclusion in the regression model.

Body Fat

- $r_{Y2|1} =$.
- $r_{Y1|2} =$.
- $r_{Y3|12} =$.

Body Fat

- $r_{Y2|1} = \sqrt{0.232} = 0.482$, since in Model 3, $\hat{\beta}_2 > 0$.
- $r_{Y1|2} = \sqrt{0.031} = 0.176$, since in Model 3, $\hat{\beta}_1 > 0$.
- $r_{Y3|12} = -\sqrt{0.105} = -0.324$, since in Model 4, $\hat{\beta}_3 < 0$.

LS Fitted Regression Coefficients as Partial Coefficients

The LS fitted regression coefficients $\hat{\beta}$ are indeed partial coefficients.

- Consider $p - 1$ X variables in the model. Let $\hat{\beta}_j$ be the LS fitted regression coefficient for X_j.
- Then $\hat{\beta}_j$ equals to the LS fitted regression coefficient when regressing the residuals $e(Y|X_{-(j)}) = Y - \hat{Y}(X_{-(j)})$ to the residuals $e(X_j|X_{-(j)}) = X_j - \hat{X}_j(X_{-(j)})$, where $X_{-(j)} = \{X_l : 1 \leq l \neq j \leq p\}$.

Confirm this with some of homework data sets.