- So far, we have always assumed that Y is linearly related to the response variables. But in many practical situations, the relationship between Y and x is nonlinear.
- If the relationship between Y and X is roughly like a curved line, then we can fit a curve (e.g. polynomial, exponential, power, logarithmic, trigonometric) instead of a straight line
- We can use the regression approach to fit these curves – a technique known as curvilinear regression

Polynomial regression

- Polynomial regression is a special case curvilinear regression
- Polynomial regression: Y is a polynomial in x
- A polynomial is any equation that has x raised to an integer power
- E.g. $Y = \beta_0 + \beta_1 x + \beta_2 x^2$ is a quadratic, producing a parabola
- E.g. $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ is a cubic in x, producing an S-shaped curve

Polynomial regression

Consider a regression model of the form

order of a polynomial
= highest exponent

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_r x_i^r + \epsilon_i,$$

r^{th} order polynomial

where

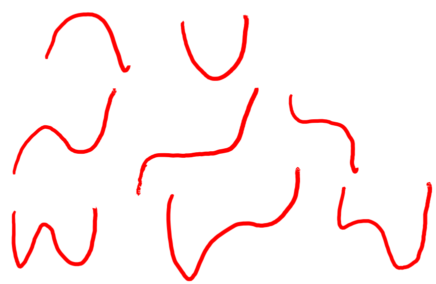
$Y_i \sim$ iid $N(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_r x_i^r, \ \sigma^2)$

$$\epsilon_i \sim i.i.d.\ N(0, \sigma^2), i = 1, 2, \ldots, n$$

e.g. $r=2$ quadratic

$r=3$ cubic

$r=4$ quartic



(U-shape)

(S-shape)

(W-shape)

Polynomial models are useful for curvilinear shapes like those above.
It is also useful for approximating unknown (and possibly complex) nonlinear relationships between Y and x, using an appropriate value of r.

Formulation

This model is not linear in x_i but it is linear in the coefficients $\beta_0, \beta_1, \dots, \beta_r$ and can be formulated as a multiple regression model by taking

$$Y = X\beta + \varepsilon$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^r \\ 1 & x_2 & x_2^2 & \dots & x_2^r \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^r \end{bmatrix} \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_r \end{bmatrix}.$$

design matrix

Least squares estimator of β is still $\hat{\beta} = (X^T X)^{-1} X^T Y$.

Remarks

- The reason this formulation is valid is that the multiple regression model makes no assumptions about the $\boldsymbol{X}$ matrix other than that its columns must be linearly independent. (It can also be proved that the $\boldsymbol{X}$ matrix will have linearly independent columns provided that r is less than the number of different x-values.)

- Prediction for polynomial regression models may be done in the same way as for general multiple regression.

- Unless there are special reasons for not doing so, whenever x^r is included in the model we should also include $x, x^2, \ldots, x^{r-1}$. hierarchy principle

Remarks

- When there is more than one predictor, multiple regression models with polynomial terms of all or some of the predictors can also be considered.

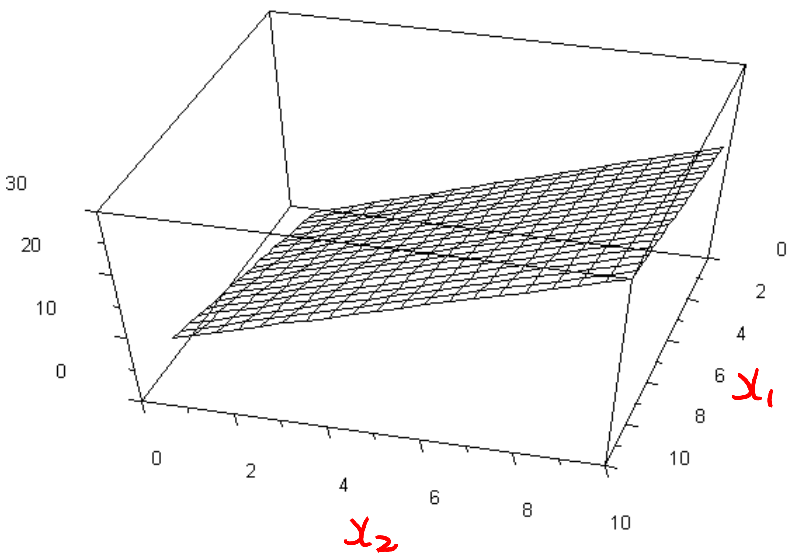
e.g. Consider two predictors x_1 and x_2. Consider $r=2$.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1}^2 + \beta_4 x_{i2}^2 + \varepsilon_i$$

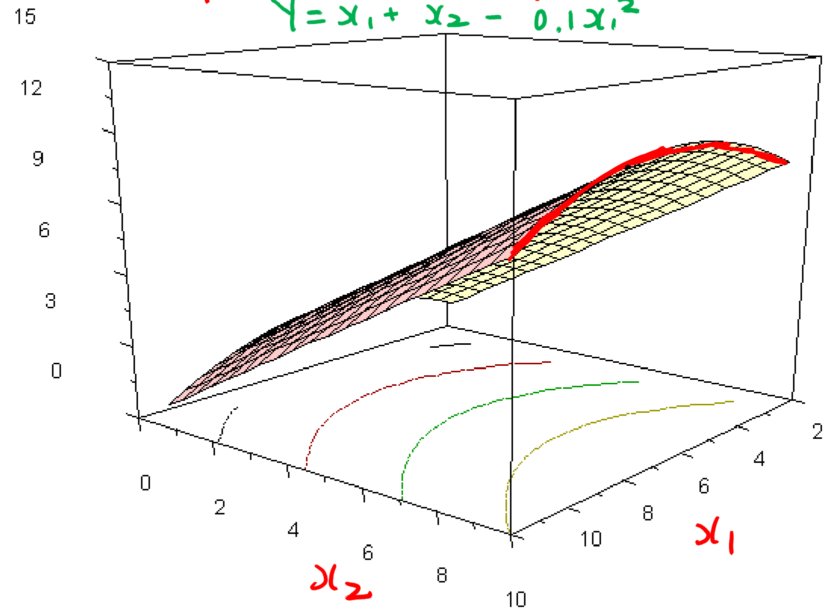
$$X = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & x_{12} & x_{12}^2 \\ 1 & x_{21} & x_{21}^2 & x_{22} & x_{22}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n1}^2 & x_{n2} & x_{n2}^2 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_4 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

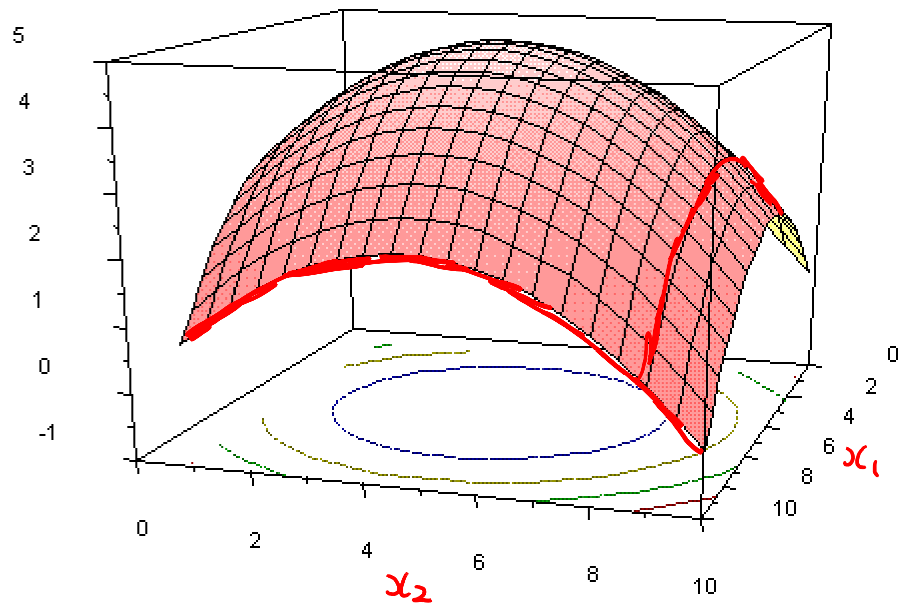
Multiple linear regression
(no polynomial terms)
$Y = x_1 + 2x_2$



One predictor with quadratic term
$Y = x_1 + x_2 - 0.1x_1^2$



Two predictors with quadratic terms



$Y = x_1 + x_2 - 0.1x_1^2 - 0.1x_2^2$

Example 4.5

The yeast data contains measurements of yield from an experiment at five different temperature levels. The variables are y=yield and x=temperature (in Fahrenheit).

x	50	50	50	70	70	70	80	80	80	90	90	90	100	100	100
y	3.3	2.8	2.9	2.3	2.6	2.1	2.5	2.9	2.4	3.0	3.1	2.8	3.3	3.5	3.0

a) Fit a linear regression to the data.

b) Fit a quadratic regression to the data.

c) Should the quadratic term be included in our model?

Example 4.5 Solution

a) $Y = \beta_0 + \beta_1 x + \varepsilon$

$$Y = \begin{bmatrix} 3.3 \\ 2.8 \\ \vdots \\ 3.0 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 50 \\ 1 & 50 \\ \vdots & \vdots \\ 1 & 100 \end{bmatrix}, \quad \hat{\beta} = (X^T X)^{-1}(X^T Y) = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 2.3063 \\ 0.0068 \end{bmatrix}.$$

b) quadratic regression $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$

Y is same as above.

$$X = \begin{bmatrix} 1 & 50 & 50^2 \\ 1 & 50 & 50^2 \\ \vdots & \vdots & \vdots \\ 1 & 100 & 100^2 \end{bmatrix}, \quad \hat{\beta} = (X^T X)^{-1}(X^T Y) = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 7.961 \\ -0.154 \\ 0.001 \end{bmatrix}.$$

c) Testing $H_0: \beta_2 = 0$ vs $H_a: \beta_2 \neq 0$.

We can do the usual t-test (for MLR).

Or we can also do an F-test comparing:

full model: $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ (quadratic model)

reduced model: $Y = \beta_0 + \beta_1 x + \varepsilon$ (linear model)

Example 4.5 Solution

```
yield <- c(3.3, 2.8, 2.9, 2.3, 2.6, 2.1, 2.5, 2.9,   2.4,
3.0, 3.1, 2.8, 3.3, 3.5, 3.0)
temperature <- c(50,  50,  50,  70,  70,  70,  80,  80,  80,
90,  90,  90, 100, 100, 100)
summary(lm(yield ~ temperature))
```

a) linear model

```
Call:
lm(formula = yield ~ temperature)

Residuals:
     Min       1Q   Median       3Q      Max
-0.67928 -0.26306  0.05315  0.22072  0.65586

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.306306   0.469075   4.917 0.000282 ***
temperature 0.006757   0.005873   1.151 0.270641
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3913 on 13 degrees of freedom
Multiple R-squared:  0.09242,   Adjusted R-squared:  0.0226
F-statistic: 1.324 on 1 and 13 DF,  p-value: 0.2706
```

Example 4.5 Solution

b)
```
summary(lm(yield ~ temperature + I(temperature^2)))
```

isolate / insulate

```
Call:
lm(formula = yield ~ temperature + I(temperature^2))

Residuals:
     Min       1Q   Median       3Q      Max
-0.37113 -0.15567 -0.04536  0.15790  0.35258

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       7.9604811  1.2589183   6.323 3.81e-05 ***
temperature      -0.1537113  0.0349408  -4.399 0.000867 ***
I(temperature^2)  0.0010756  0.0002329   4.618 0.000592 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2444 on 12 degrees of freedom
Multiple R-squared:  0.6732,    Adjusted R-squared:  0.6187
F-statistic: 12.36 on 2 and 12 DF,  p-value: 0.001218
```

c) t-test of $H_0: \beta_2 = 0$

test statistic: $t = 4.618$

P-value: 0.000592

Using $\alpha = 0.05$, we reject H_0.

We should keep the quadratic term in our model.

Alternative ways of fitting this model in R:

① temperature2 <- temperature^2; lm(yield ~ temperature + temperature2)

② summary(lm(yield ~ poly(temperature, 2, raw=T)))

Example 4.5 Solutions (cont.)

c)
```
full <- lm(yield ~ temperature + I(temperature^2))
reduced <- lm(yield ~ temperature)
anova(reduced, full)
```

```
Analysis of Variance Table

Model 1: yield ~ temperature
Model 2: yield ~ temperature + I(temperature^2)
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1     13 1.99063
2     12 0.71677  1    1.2739 21.327 0.0005921 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The P-value is significant (at 5% level). Hence there is sufficient evidence to reject $H_0: \beta_2 = 0$. The quadratic term should be kept in our model.

Note that the R output from anova() is displayed differently to our ANOVA table. Try working out the correspondence between the two as an exercise.

Example 4.6

	x_1	x_2	x_3
Odor	**Temperature**	**Ratio**	**Height**
66	-1	-1	0
58	-1	0	-1
65	0	-1	-1
-31	0	0	0
39	1	-1	0
17	1	0	-1
7	0	1	-1
-35	0	0	0
43	-1	1	0
-5	-1	0	1
43	0	-1	1
-26	0	0	0
49	1	1	0
-40	1	0	1
-22	0	1	1

An experiment was designed to relate three variables (temperature, ratio, and height) to a measure of odor in a chemical process. Each variable has 3 levels, but the design was not constructed as a full factorial design.

Fit a polynomial regression to the data, considering all first and second order terms.

$x_1, x_1^2, x_2, x_2^2, x_3, x_3^2$

Example 4.6 Solution

```
Temp <- c(-1,-1,0,0,1,1,0,0,-1,-1,0,0,1,1,0)
Ratio <- c(-1,0,-1,0,-1,0,1,0,1,0,-1,0,1,0,1)
Height <-c(0,-1,-1,0,0,-1,-1,0,0,1,1,0,0,1,1)
Odor <- c(66,58,65,-31,39,17,7,-35,43,-5,43,-26,49,-40,-22)
odor <- data.frame(Odor,Temp,Ratio,Height)
odor
```

```
   Odor Temp Ratio Height
1    66   -1    -1      0
2    58   -1     0     -1
3    65    0    -1     -1
4   -31    0     0      0
5    39    1    -1      0
6    17    1     0     -1
7     7    0     1     -1
8   -35    0     0      0
9    43   -1     1      0
10   -5   -1     0      1
11   43    0    -1      1
12  -26    0     0      0
13   49    1     1      0
14  -40    1     0      1
15  -22    0     1      1
```

Example 4.6 Solution

```
summary(lm(Odor ~
Temp+Ratio+Height+I(Temp^2)+I(Ratio^2)+I(Height^2), data=odor))
```

```
Call:
lm(formula = Odor ~ Temp + Ratio + Height + I(Temp^2) + I(Ratio^2) +
    I(Height^2), data = odor)

Residuals:
    Min      1Q  Median      3Q     Max
-20.625  -9.625  -1.375   4.021  28.875

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -30.667     10.840  -2.829   0.0222 *
Temp         -12.125      6.638  -1.827   0.1052
Ratio        -17.000      6.638  -2.561   0.0336 *
Height       -21.375      6.638  -3.220   0.0122 *
I(Temp^2)     32.083      9.771   3.284   0.0111 *
I(Ratio^2)    47.833      9.771   4.896   0.0012 **
I(Height^2)    6.083      9.771   0.623   0.5509
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.77 on 8 degrees of freedom
Multiple R-squared:  0.8683,    Adjusted R-squared:  0.7695
F-statistic: 8.789 on 6 and 8 DF,  p-value: 0.003616
```

Use backward selection.
Start with the full model.
Find the largest P-value
that is greater than 0.05
(our level of significance).

drop $Height^2$

Example 4.6 Solution

```
summary(lm(Odor ~ Temp+Ratio+Height+I(Temp^2)+I(Ratio^2),
data=odor))
```

```
Call:
lm(formula = Odor ~ Temp + Ratio + Height + I(Temp^2) + I(Ratio^2),
    data = odor)

Residuals:
    Min      1Q  Median      3Q     Max
-17.933  -9.635  -4.067   4.620  26.933

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -26.923      8.707  -3.092 0.012884 *
Temp         -12.125      6.408  -1.892 0.091024 .
Ratio        -17.000      6.408  -2.653 0.026350 *
Height       -21.375      6.408  -3.336 0.008720 **
I(Temp^2)     31.615      9.404   3.362 0.008366 **
I(Ratio^2)    47.365      9.404   5.036 0.000703 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.12 on 9 degrees of freedom
Multiple R-squared:  0.8619,    Adjusted R-squared:  0.7852
F-statistic: 11.23 on 5 and 9 DF,  p-value: 0.001169
```

cannot drop Temp, as $Temp^2$ is significant.

So this becomes our final model.

Interaction terms

Consider a model with two predictors x_{i1} and x_{i2}. We can also consider an interaction between these predictors:

interaction term

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta x_{i2} + \beta_3 x_{i1} x_{i2} + \epsilon_i,$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}x_{n2} \end{bmatrix}$$

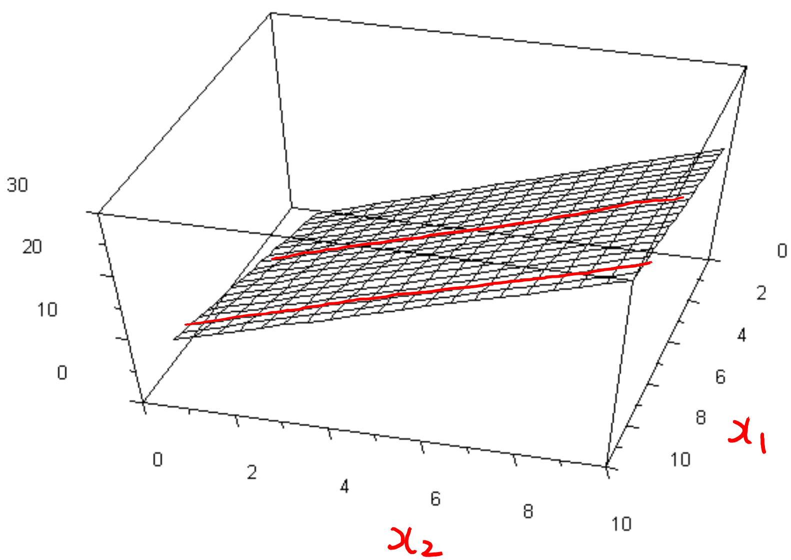
where

$$\epsilon_i \sim i.i.d.\, N(0, \sigma^2), i = 1, 2, \ldots, n$$

Interaction terms allow us to model the relationship between the response and predictors when one predictor will vary with changes in value of another predictor.

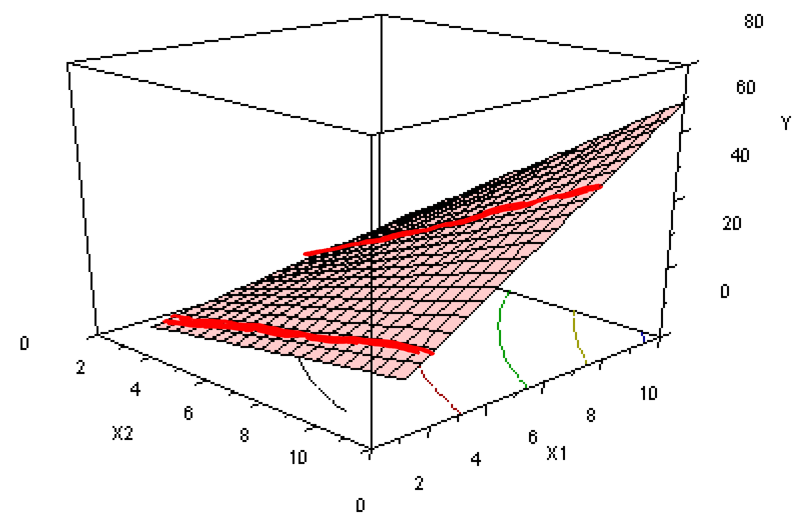
This is another way that will produce curved shapes.

no interaction



slope of x_2 is the same for different values of x_1 (and vice versa)

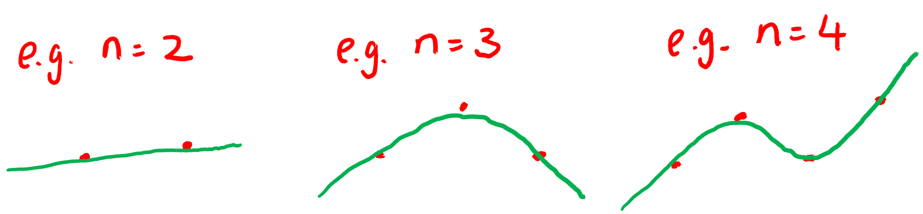
with interaction

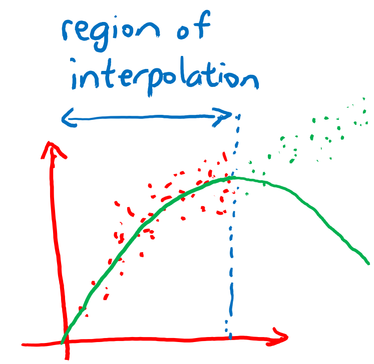


slope of x_2 changes with different values of x_1 (and vice versa)

Considerations when fitting polynomials

1) Order of the polynomial
   - This should be kept as low as possible
   - Use a model that is consistent with the knowledge of the data and the context
   - Fitting an arbitrarily high polynomial can lead to overfitting. It is always possible to fit a polynomial of order $(n-1)$ that will pass through all n points.





2) Extrapolation
   - Curvature in the region of the data and in the region outside of the data can be different. So prediction may be inaccurate.

Considerations when fitting polynomials

3) Ill-conditioning
   - Recall one of the assumptions in linear regression is that the columns of X must be linearly independent. It should have full rank.
   - In polynomial regression, as the order increases, $X^\top X$ will start to become ill-conditioned. Hence, $(X^\top X)^{-1}$ may not be accurate, and so $\hat{\beta}$ may be incorrect.
   - If x lies in a narrow range, then the degree of ill-conditioning increases. Also, multicolinearity will start entering in.