Regression extensions

Chrysafis Vogiatzis

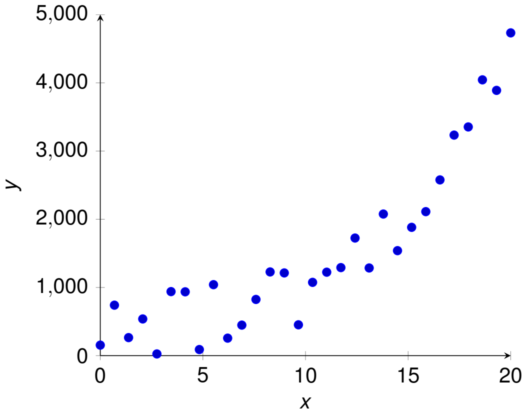
Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 33

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

Polynomial regression

What if our data looks like this?



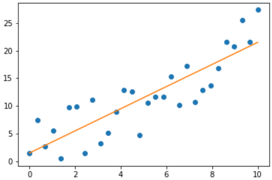
Polynomial regression

Which one appears to "fit best"?

Polynomial regression

Which one appears to "fit best"?

$$\hat{y} = \beta_0 + \beta_1 x$$

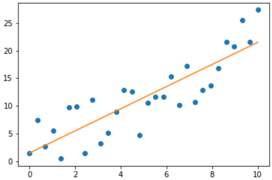


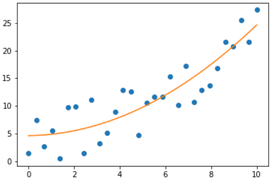
Polynomial regression

Which one appears to “fit best”?

$$\hat{y} = \beta_0 + \beta_1 x$$

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$

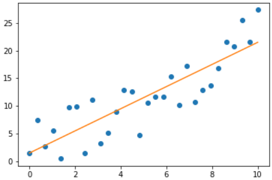




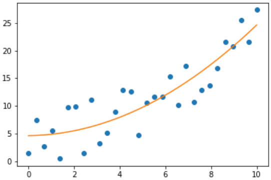
Polynomial regression

Which one appears to “fit best”?

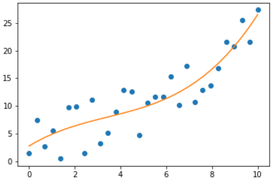
$$\hat{y} = \beta_0 + \beta_1 x$$



$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$



$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

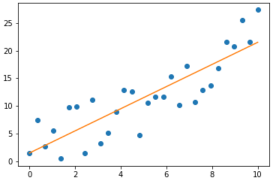


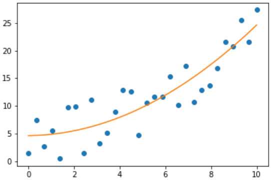
Which one appears to "fit best"?

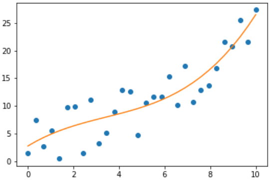
$$\hat{y} = \beta_0 + \beta_1 x$$

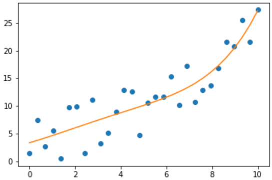
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$$

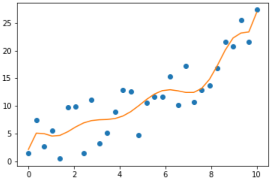
$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$











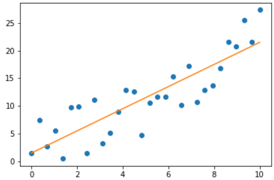
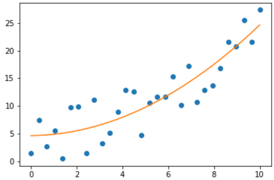
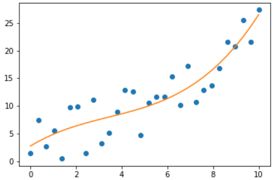
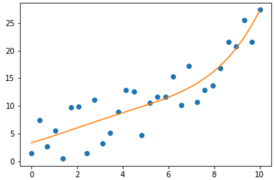
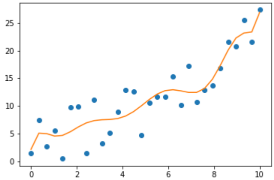
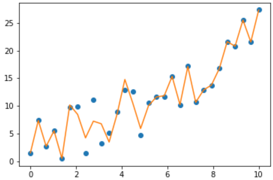
Polynomial regression

Which one appears to "fit best"?

$\hat{y} = \beta_0 + \beta_1 x$

$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2$

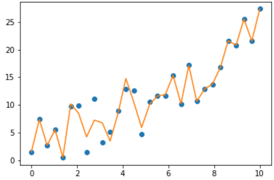
$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$

Appropriate model selection

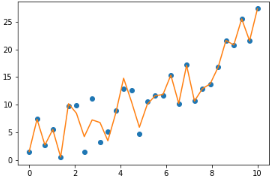
Appropriate model selection

Overfitting

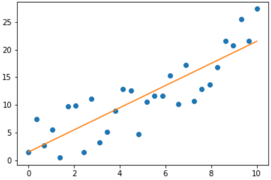


Appropriate model selection

Overfitting

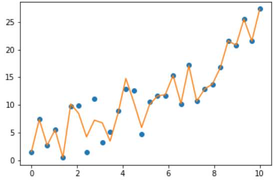


Underfitting

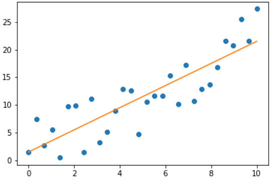


Appropriate model selection

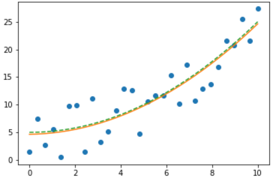
Overfitting



Underfitting



Appropriate model



Polynomial regression: a small example

How to fit data points with a line of the form:

$$y = \beta_0 + \beta_1 x + \beta_{11} x_1^2?$$

Pretty simple idea!

- First, create a “new” predictor variable x_2.
- Set it equal to x_1^2!
- Create matrix X based on x_1 and $x_2 = x_1^2$.
- Solve for $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_{11} \end{bmatrix} = \left(X^T X\right)^{-1} X^T y$.

Polynomial regression: a small example

How to fit data points with a line of the form:

$$y = \beta_0 + \beta_1 x + \beta_{11} x_1^2?$$

Pretty simple idea!

- First, create a "new" predictor variable x_2.
- Set it equal to x_1^2!
- Create matrix X based on x_1 and $x_2 = x_1^2$.
- Solve for $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_{11} \end{bmatrix} = \left(X^T X\right)^{-1} X^T y$.

ILLINOIS

Polynomial regression: a small example

How to fit data points with a line of the form:

$$y = \beta_0 + \beta_1 x + \beta_{11} x_1^2?$$

Pretty simple idea!

- First, create a "new" predictor variable x_2.
- Set it equal to x_1^2!
- Create matrix X based on x_1 and $x_2 = x_1^2$.
- Solve for $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_{11} \end{bmatrix} = (X^T X)^{-1} X^T y$.

Polynomial regression: a small example

How to fit data points with a line of the form:

$$y = \beta_0 + \beta_1 x + \beta_{11} x_1^2?$$

Pretty simple idea!

- First, create a "new" predictor variable x_2.
- Set it equal to x_1^2!
- Create matrix X based on x_1 and $x_2 = x_1^2$.

- Solve for $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_{11} \end{bmatrix} = \left(X^T X\right)^{-1} X^T y.$

Polynomial regression: a small example

How to fit data points with a line of the form:

$$y = \beta_0 + \beta_1 x + \beta_{11} x_1^2?$$

Pretty simple idea!

- First, create a “new” predictor variable x_2.
- Set it equal to x_1^2!
- Create matrix X based on x_1 and $x_2 = x_1^2$.
- Solve for $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_{11} \end{bmatrix} = \left(X^T X\right)^{-1} X^T y$.

ILLINOIS

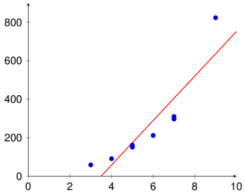
Example

Consider the following data:

x	y
7	310
3	59
5	153
5	162
4	91
6	212
7	297
5	151
9	823

We tried a linear regression and got the line:

$$y = 7.2404x - 2.2194.$$



Since it does not look great, we decide to try a second degree polynomial regression of the form: $y = \beta_0 + \beta_1 x + \beta_{11} x^2$.

Example solution

1. First, create a new column in the data: x^2.

x	x^2	y
7	49	310
3	9	59
5	25	153
5	25	162
4	16	91
6	36	212
7	49	297
5	25	151
9	81	823

2. Build X.

$$X = \begin{bmatrix} 1 & 7 & 49 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 5 & 25 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 5 & 25 \\ 1 & 9 & 81 \end{bmatrix}$$

Example solution

1. First, create a new column in the data: x^2.

x	x^2	y
7	49	310
3	9	59
5	25	153
5	25	162
4	16	91
6	36	212
7	49	297
5	25	151
9	81	823

2. Build X.

Example solution

1. First, create a new column in the data: x^2.

x	x^2	y
7	49	310
3	9	59
5	25	153
5	25	162
4	16	91
6	36	212
7	49	297
5	25	151
9	81	823

2. Build X.

$$X = \begin{bmatrix} 1 & 7 & 49 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 5 & 25 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 5 & 25 \\ 1 & 9 & 81 \end{bmatrix}$$

ILLINOIS

Example solution

1. First, create a new column in the data: x^2.

x	x^2	y
7	49	310
3	9	59
5	25	153
5	25	162
4	16	91
6	36	212
7	49	297
5	25	151
9	81	823

2. Build X.

$$X = \begin{bmatrix} 1 & 7 & 49 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 5 & 25 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 5 & 25 \\ 1 & 9 & 81 \end{bmatrix}$$

ILLINOIS

Example solution

1. First, create a new column in the data: x^2.

x	x^2	y
7	49	310
3	9	59
5	25	153
5	25	162
4	16	91
6	36	212
7	49	297
5	25	151
9	81	823

2. Build X.

$$X = \begin{bmatrix} 1 & 7 & 49 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 5 & 25 \\ 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 5 & 25 \\ 1 & 9 & 81 \end{bmatrix}$$

ILLINOIS

Example solution

3. Solve for $\hat{\beta}$:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_{11} \end{bmatrix} = \left(X^T X\right)^{-1} X^T y = \begin{bmatrix} 437.74 \\ -190.47 \\ 25.5 \end{bmatrix}$$

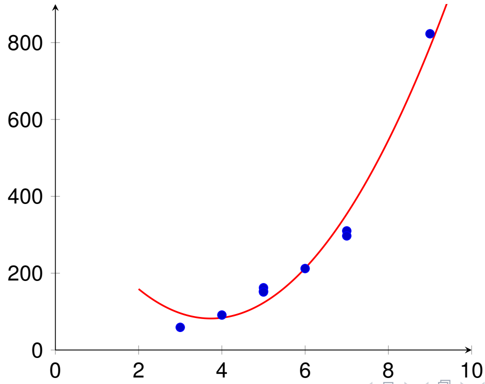
4. Plot $y = 437.74 - 190.47x_1 + 25.5x_1^2$:

Example solution

3 Solve for $\hat{\beta}$:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_{11} \end{bmatrix} = \left(X^T X\right)^{-1} X^T y = \begin{bmatrix} 437.74 \\ -190.47 \\ 25.5 \end{bmatrix}$$

4 Plot $y = 437.74 - 190.47x_1 + 25.5x_1^2$:



Other interactions

We can use that same logic for other, different variable interactions and functions. For example:

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$
  - Introduce new variable $x_{12} = x_1 x_2$ and solve.
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{123} x_1 x_2 x_3$
  - Introduce new variable $x_{123} = x_1 x_2 x_3$ and solve.
- We can even do that with other nonlinear functions: for example $y = \beta_0 + \beta_1 x_1 + \beta_2 cos(x_1)$.
  - Introduce new variable $x_2 = cos(x_1)$ and solve.
- Or $y = \beta_0 + \beta_1 x_1 + \beta_2 \log x_1$.
  - Introduce new variable $x_2 = \log x_1$ and solve.

Other interactions

We can use that same logic for other, different variable interactions and functions. For example:

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$
  - Introduce new variable $x_{12} = x_1 x_2$ and solve.
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{123} x_1 x_2 x_3$
  - Introduce new variable $x_{123} = x_1 x_2 x_3$ and solve.
- We can even do that with other nonlinear functions: for example $y = \beta_0 + \beta_1 x_1 + \beta_2 cos(x_1)$.
  - Introduce new variable $x_2 = cos(x_1)$ and solve.
- Or $y = \beta_0 + \beta_1 x_1 + \beta_2 \log x_1$.
  - Introduce new variable $x_2 = \log x_1$ and solve.

ILLINOIS

Other interactions

We can use that same logic for other, different variable interactions and functions. For example:

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$
  - Introduce new variable $x_{12} = x_1 x_2$ and solve.
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{123} x_1 x_2 x_3$
  - Introduce new variable $x_{123} = x_1 x_2 x_3$ and solve.
- We can even do that with other nonlinear functions: for example $y = \beta_0 + \beta_1 x_1 + \beta_2 cos(x_1)$.
  - Introduce new variable $x_2 = cos(x_1)$ and solve.
- Or $y = \beta_0 + \beta_1 x_1 + \beta_2 \log x_1$.
  - Introduce new variable $x_2 = \log x_1$ and solve.

Other interactions

We can use that same logic for other, different variable interactions and functions. For example:

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$
  - Introduce new variable $x_{12} = x_1 x_2$ and solve.
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{123} x_1 x_2 x_3$
  - Introduce new variable $x_{123} = x_1 x_2 x_3$ and solve.
- We can even do that with other nonlinear functions: for example $y = \beta_0 + \beta_1 x_1 + \beta_2 cos(x_1)$.
  - Introduce new variable $x_2 = cos(x_1)$ and solve.
- Or $y = \beta_0 + \beta_1 x_1 + \beta_2 \log x_1$.
  - Introduce new variable $x_2 = \log x_1$ and solve.

ILLINOIS

Model selection

Given k predictor variables, we saw that not all need to be significant. So, this begs the question: which variables should I include in my regression?

1. **All subsets selection.**
   - Consider all (2^k) possible combinations of variables.
   - Pick the model with highest R^2_{adj}.
2. **Backwards selection.**
3. **Forwards selection.**

Model selection

Given k predictor variables, we saw that not all need to be significant. So, this begs the question: which variables should I include in my regression?

1. **All subsets selection.**
   - Consider all (2^k) possible combinations of variables.
   - Pick the model with highest R^2_{adj}.
2. **Backwards selection.**
   - Start from a regression including all variables.
   - Keep removing the least significant variable until R^2_{adj} starts decreasing.
3. **Forwards selection.**
   - Start from a regression including no variables.
   - Keep adding the most significant variable until R^2_{adj} starts decreasing.

Model selection

Given k predictor variables, we saw that not all need to be significant. So, this begs the question: which variables should I include in my regression?

1. **All subsets selection.**
   - Consider all (2^k) possible combinations of variables.
   - Pick the model with highest R^2_{adj}.
2. **Backwards selection.**
   - Start from a regression including all variables.
   - Keep removing the least significant variable until R^2_{adj} starts decreasing.
3. **Forwards selection.**
   - Start from a regression including no variables.
   - Keep adding the most significant variable until R^2_{adj} starts decreasing.

Model selection

Given k predictor variables, we saw that not all need to be significant. So, this begs the question: which variables should I include in my regression?

1. **All subsets selection.**
   - Consider all (2^k) possible combinations of variables.
   - Pick the model with highest R^2_{adj}.
2. **Backwards selection.**
   - Start from a regression including all variables.
   - Keep removing the least significant variable until R^2_{adj} starts decreasing.
3. **Forwards selection.**
   - Start from a regression including no variables.
   - Keep adding the most significant variable until R^2_{adj} starts decreasing.

ILLINOIS

Model selection

Given k predictor variables, we saw that not all need to be significant. So, this begs the question: which variables should I include in my regression?

1. **All subsets selection.**
   - Consider all (2^k) possible combinations of variables.
   - Pick the model with highest R^2_{adj}.
2. **Backwards selection.**
   - Start from a regression including all variables.
   - Keep removing the least significant variable until R^2_{adj} starts decreasing.
3. **Forwards selection.**
   - Start from a regression including no variables.
   - Keep adding the most significant variable until R^2_{adj} starts decreasing.

Validating a regression model

How can we validate our model?

- Split our data into two parts:
  - training data
  - testing data
- Common split is 80%-20% (in favor of training).
- Use the training data to create the regression.
- Use the testing data to test how well the regression is performing.
- Check the performance by calculating the MS_E:

$$MS_E = \frac{1}{n-2}\sum\left(y_i^{test} - \hat{y}_i^{test}\right)^2.$$

ILLINOIS

Validating a regression model

How can we validate our model?

- Split our data into two parts:
  - training data
  - testing data
- Common split is 80%-20% (in favor of training).
- Use the training data to create the regression.
- Use the testing data to test how well the regression is performing.
- Check the performance by calculating the MS_E:

$$MS_E = \frac{1}{n-2} \sum \left(y_i^{test} - \hat{y}_i^{test} \right)^2 .$$

Validating a regression model

How can we validate our model?

- Split our data into two parts:
  - training data
  - testing data
- Common split is 80%-20% (in favor of training).
- Use the training data to create the regression.
- Use the testing data to test how well the regression is performing.
- Check the performance by calculating the MS_E:

$$MS_E = \frac{1}{n-2} \sum \left(y_i^{test} - \hat{y}_i^{test} \right)^2 .$$

Validating a regression model

How can we validate our model?

- Split our data into two parts:
  - training data
  - testing data
- Common split is 80%-20% (in favor of training).
- Use the training data to create the regression.
- Use the testing data to test how well the regression is performing.
- Check the performance by calculating the MS_E:

$$MS_E = \frac{1}{n-2} \sum \left(y_i^{test} - \hat{y}_i^{test} \right)^2 .$$

Validating a regression model

How can we validate our model?

- Split our data into two parts:
  - training data
  - testing data
- Common split is 80%-20% (in favor of training).
- Use the training data to create the regression.
- Use the testing data to test how well the regression is performing.
- Check the performance by calculating the MS_E:

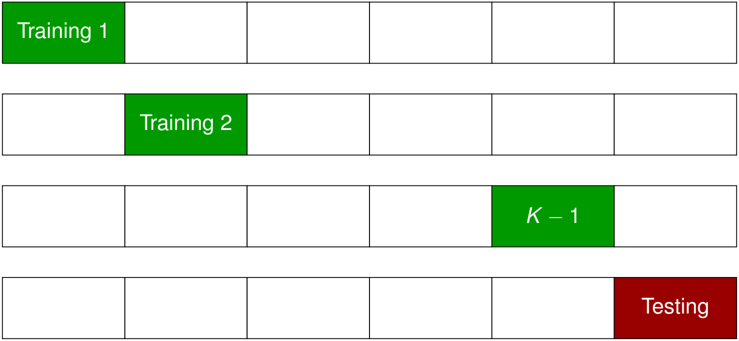
$$MS_E = \frac{1}{n-2} \sum \left(y_i^{test} - \hat{y}_i^{test}\right)^2.$$

K-fold validation

- Split our data into K parts:
  - $K - 1$ parts with training data.
  - 1 part of testing data.
- Use the training data to create $K - 1$ models.
- Use the testing data to test how well **each of the regressions** are performing.
- Output the best model amongst them.

K-fold validation

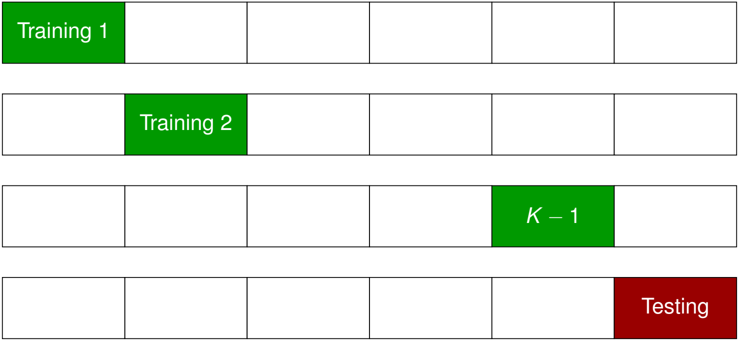
- Split our data into K parts:
  - $K - 1$ parts with training data.
  - 1 part of testing data.
- Use the training data to create $K - 1$ models.
- Use the testing data to test how well **each of the regressions** are performing.
- Output the best model amongst them.



ILLINOIS

K-fold validation

- Split our data into K parts:
  - $K-1$ parts with training data.
  - 1 part of testing data.
- Use the training data to create $K-1$ models.
- Use the testing data to test how well **each of the regressions** are performing.
- Output the best model amongst them.



K-fold validation

- Split our data into K parts:
  - $K - 1$ parts with training data.
  - 1 part of testing data.
- Use the training data to create $K - 1$ models.
- Use the testing data to test how well **each of the regressions** are performing.
- Output the best model amongst them.