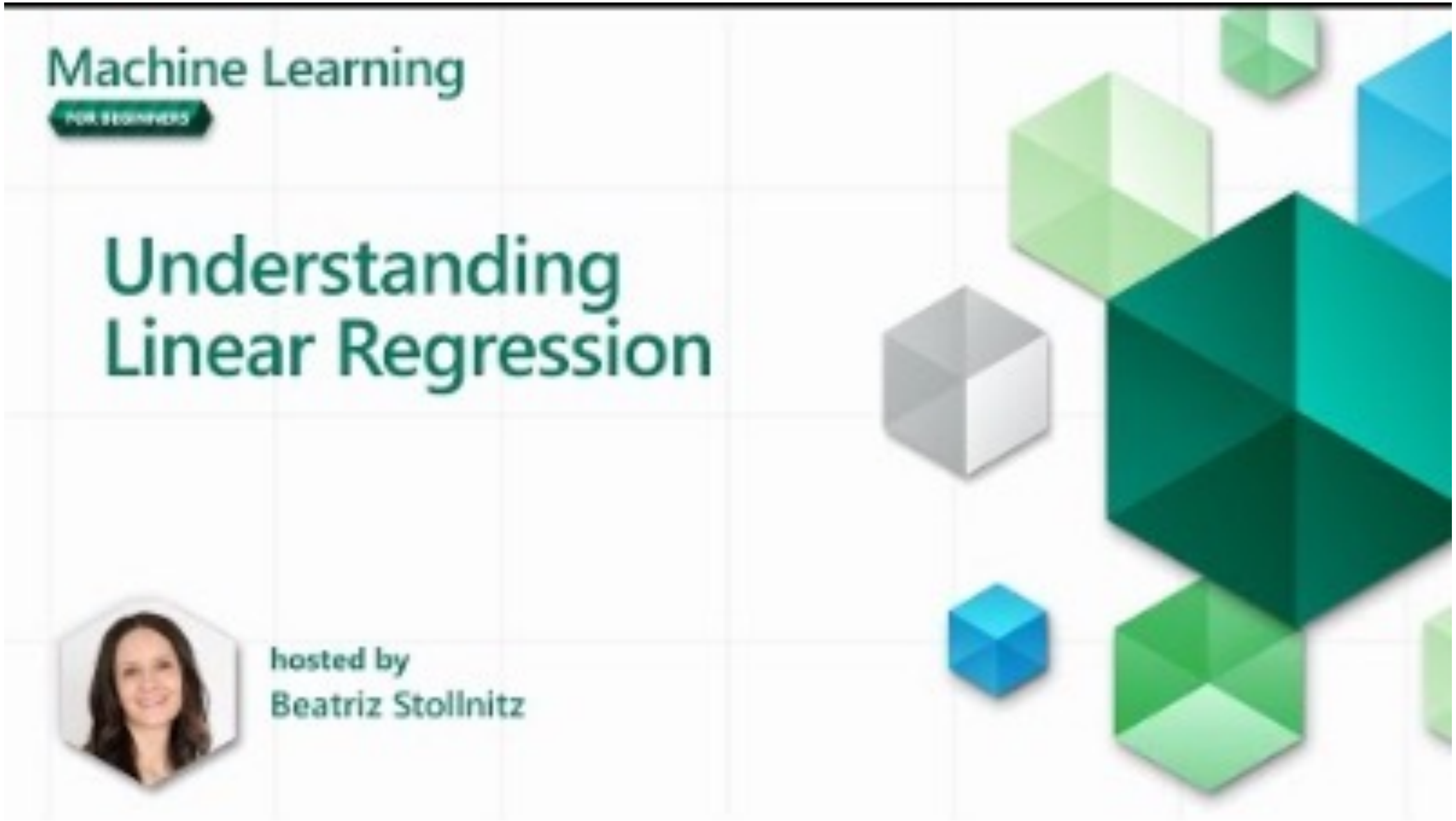
ARTIFICIAL INTELLIGENCE NEEDS REAL INTELLIGENCE

Regression II

Professor: Tianyuan ZHANG
tianyuan.zhang@kedgebs.com

KEDGE BUSINESS SCHOOL

Recap of Previous Session

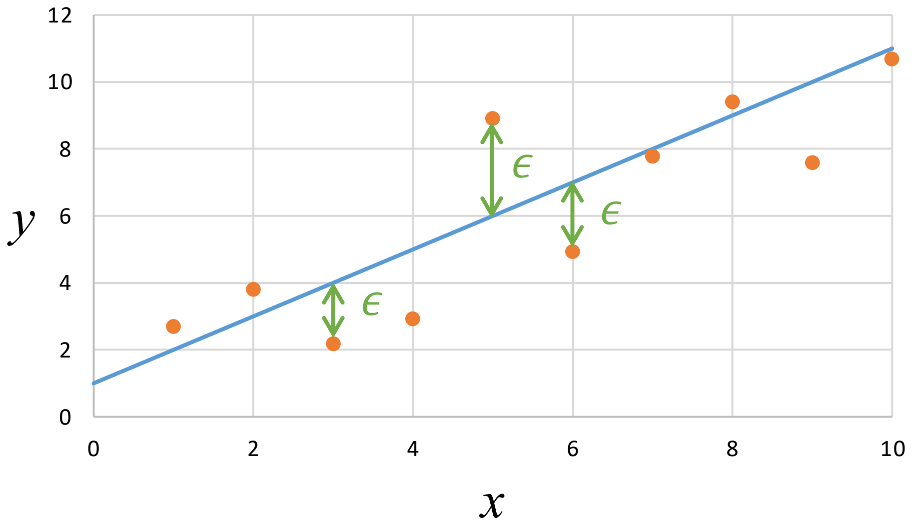


Recap of Previous Session

- Regression
  - Supervised learning
  - Learn the relationships between features $(x_1, x_2, \ldots, x_n)$ and label y
  - Use the learned relationships to make prediction $\hat{y}$ on new data
- Linear regression
  - Estimate the linear relationship between $(x_1, x_2, \ldots, x_n)$ and y
  - $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_N x_N$
- Simple linear regression
  - A single input feature
  - $y = \beta_0 + \beta_1 x$

Recap of Previous Session

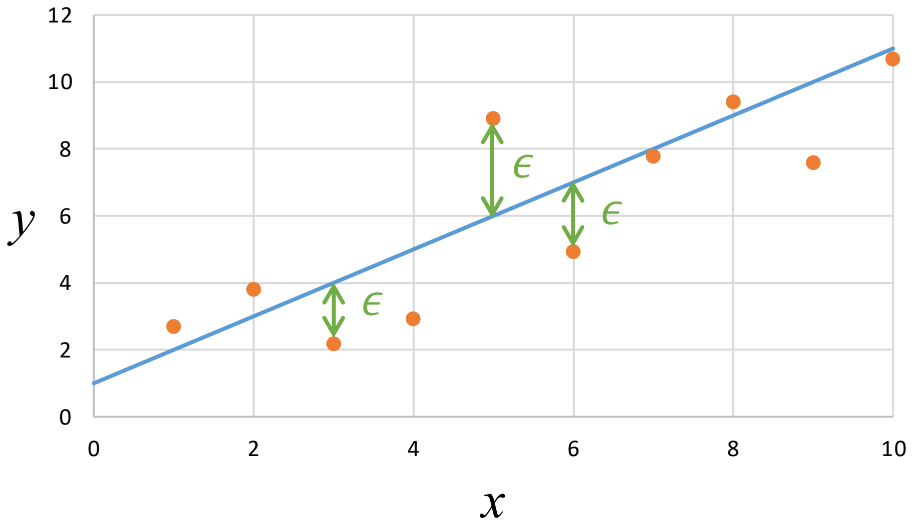
- Simple linear regression
  - $y = \beta_0 + \beta_1 x$
  - For (x_i, y_i), the prediction $\hat{y}_i$
    - The prediction $\hat{y}_i = \beta_0 + \beta_1 x_i$
    - The residual $\epsilon_i = y_i - \hat{y}_i$
  - The goal is to find the line that best fits the data points
    - Find the best value of β_0 and β_1
    - Minimize the residuals for all data points



KEDGE BUSINESS SCHOOL

Recap of Previous Session

- Simple linear regression
  - $y = \beta_0 + \beta_1 x$
  - The goal is to find the line that best fits the data points
    - Find the best value of β_0 and β_1
    - Minimize the residuals for all data points
- Ordinary least squares
  - An approach to estimate β_0 and β_1 by minimizing the sum of squared errors
    - Minimize $\sum_i \epsilon_i^2 = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2$
  - The estimated $\hat{\beta}_0$ and $\hat{\beta}_1$



Recap of Previous Session

- Evaluation metrics for regression
  - Measure the model performance on unseen data
  - Residual, $\epsilon_i = y_i - \hat{y}_i$
  - Mean squared error, $\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n} \epsilon_i^2$
  - Root mean squared error, $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$
  - Mean absolute error, $\mathrm{MAE} = \frac{1}{n}\sum_{i=0}^{n} |\epsilon_i|$
  - Coefficient of determination, $R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \bar{y})^2}$, $\bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$

KEDGE BUSINESS SCHOOL

Recap of Previous Session

- Evaluation metrics for regression
  - Measure the model performance on unseen data
  - Residual, $\epsilon_i = y_i - \hat{y}_i$
  - Mean squared error, $\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n} \epsilon_i^2$
  - Root mean squared error, $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$
  - Mean absolute error, $\mathrm{MAE} = \frac{1}{n}\sum_{i=0}^{n} |\epsilon_i|$
  - Coefficient of determination, $R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \bar{y})^2}$, $\bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$

The lower values are better.

Recap of Previous Session

- Evaluation metrics for regression
  - Measure the model performance on unseen data
  - Residual, $\epsilon_i = y_i - \hat{y}_i$
  - Mean squared error, $\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n} \epsilon_i^2$
  - Root mean squared error, $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$
  - Mean absolute error, $\mathrm{MAE} = \frac{1}{n}\sum_{i=0}^{n}|\epsilon_i|$
  - Coefficient of determination, $R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \bar{y})^2}$, $\bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$

1. *Close to 1 → Good fitness*
2. *Equal to 0 → Model returns the average value of y*
3. *Less than 0 → Model learned wrong relationships*

Recap of Previous Session

- Evaluation metrics for regression
  - Measure the model performance on unseen data
  - Residual, $\epsilon_i = y_i - \hat{y}_i$
  - Mean squared error, $\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n} \epsilon_i^2$
  - Root mean squared error, $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$
  - Mean absolute error, $\mathrm{MAE} = \frac{1}{n}\sum_{i=0}^{n} |\epsilon_i|$
  - Coefficient of determination, $R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n}(y_i - \bar{y})^2}$, $\bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$

All affected by outliers

Recap of Previous Session

- Evaluation metrics for regression
  - Measure the model performance on unseen data
  - Residual, $\epsilon_i = y_i - \hat{y}_i$
  - Mean squared error, $\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n} \epsilon_i^2$
  - Root mean squared error, $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$
  - Mean absolute error, $\mathrm{MAE} = \frac{1}{n}\sum_{i=0}^{n}|\epsilon_i|$

    Less affected by outliers
  - Coefficient of determination, $R^2 = 1 - \frac{\sum_{i=0}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=0}^{n}(y_i-\bar{y})^2}$, $\bar{y} = \frac{1}{n}\sum_{i=0}^{n} y_i$

KEDGE BUSINESS SCHOOL

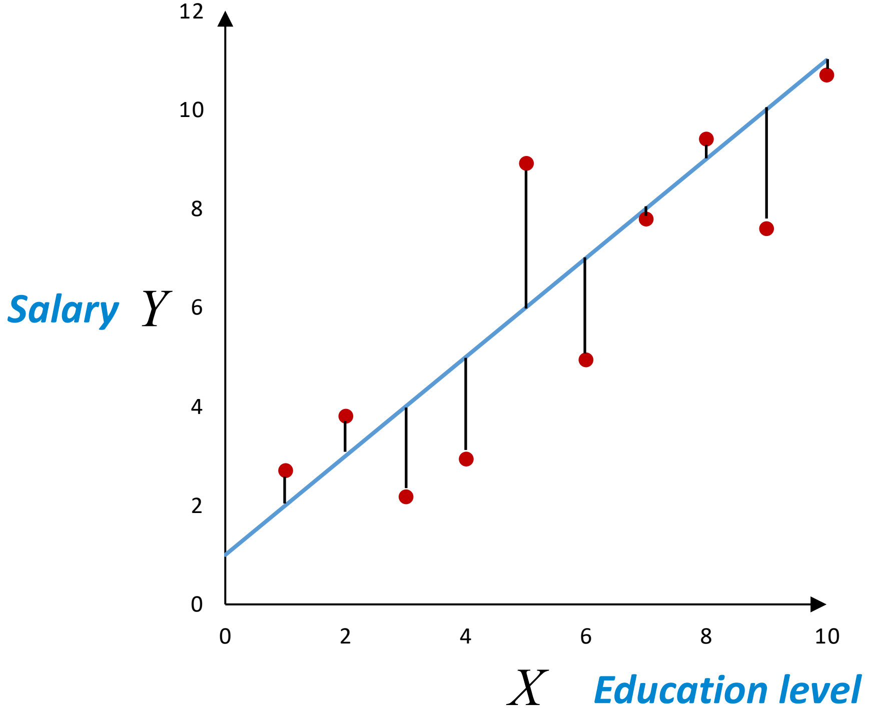
Outline

- **Multiple Linear Regression**
- Feature Selection
- Polynomial Regression
- Under-fitting & Over-fitting

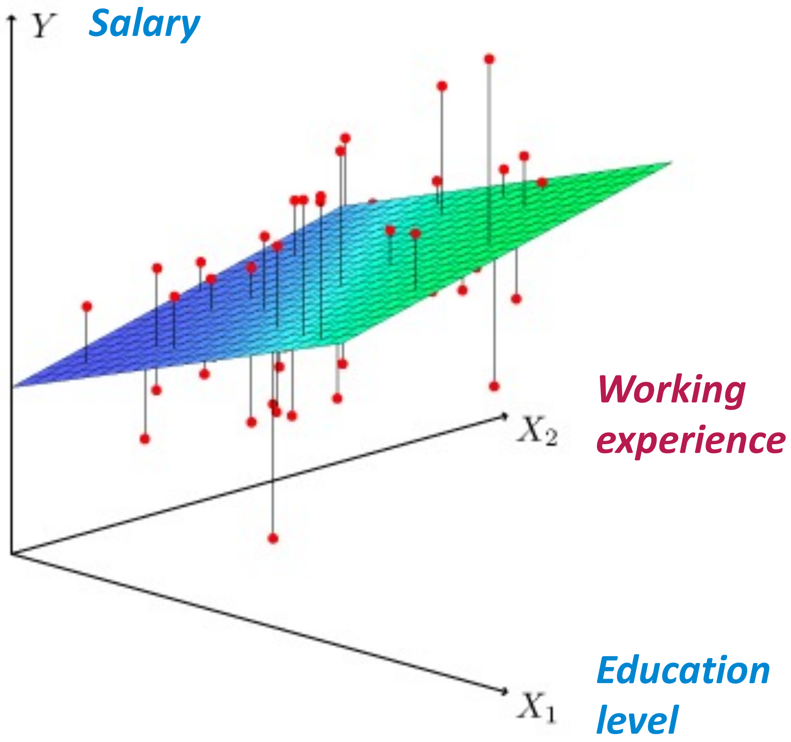
Multiple Linear Regression

- Simple linear regression
  - Linear regression with a single independent variable x
  - $y = \beta_0 + \beta_1 x$
- Multiple linear regression
  - Linear regression with more than one independent variables $x_1, x_2, \dots, x_n$
  - $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_N x_N$
  - The general form of linear regression

Multiple Linear Regression



Simple linear regression



Multiple linear regression with 2 independent variables

Image source: https://medium.com/analytics-vidhya/multiple-linear-regression-an-intuitive-approach-f874f7a6a7f9

Multiple Linear Regression

- Multiple linear regression
  - $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_N x_N$
  - Make predictions based on multiple factors
  - Estimate a plane in the 3-dimensional space when there are 2 independent variables
  - Estimate a hyper-plane in the $(n + 1)$-dimensional space when there are n independent variables

Multiple Linear Regression

- `sklearn.linear_model.LinearRegression`
  - For both simple and multiple linear regression

fit(*X*, *y*, *sample_weight=None*) [source]

Fit linear model.

Parameters:	**X : *{array-like, sparse matrix} of shape (n_samples, n_features)*** Training data. **y : *array-like of shape (n_samples,) or (n_samples, n_targets)*** Target values. Will be cast to X's dtype if necessary. **sample_weight : *array-like of shape (n_samples,), default=None*** Individual weights for each sample. *New in version 0.17:* parameter *sample_weight* support to LinearRegression.
Returns:	**self : *object*** Fitted Estimator.

Number of features

Multiple Linear Regression

- `sklearn.linear_model.LinearRegression`

```
# if there is only a single feature
X_train.shape
```

```
(221,)
```

```
# fit the model
slr = LinearRegression(fit_intercept = True)
slr.fit(X_train, y_train)
```

```
ValueError: Expected 2D array, got 1D array instead:
```

```
X_train.reshape(-1, 1).shape
```

```
(221, 1)
```

Multiple Linear Regression

- `sklearn.linear_model.LinearRegression`

```
# if there are more then one features
X_train.shape
```

```
(221, 10)
```

```
# fit the model
mlr = LinearRegression(fit_intercept = True)
mlr.fit(X_train, y_train)
```

```
▾ LinearRegression
LinearRegression()
```

Outline

- Multiple Linear Regression
- **Feature Selection**
- Polynomial Regression
- Under-fitting & Over-fitting

Feature Selection

- Definition
  - The process of selecting a subset of relevant features for model construction
- Objectives
  - Improve model performance by selecting relevant features and removing irrelevant ones
  - Simplify the model to make it easier to interpret
  - Reduce training time by reducing the amount of computation
  - Enhance model applicability by reducing the number of required features

Feature Selection

- Definition
  - The process of selecting a subset of relevant features for model construction
- Based on experience

Location	Size	~~Name of previous owner~~	Construction year	~~Post code~~	House price
Available features					*Label / Target*

Feature Selection

- Definition
  - The process of selecting a subset of relevant features for model construction
- Based on quantitative indicators

	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6	target
age	1.000000	0.173737	0.185085	0.335428	0.260061	0.219243	-0.075181	0.203841	0.270774	0.301731	0.187889
sex	0.173737	1.000000	0.088161	0.241010	0.035277	0.142637	-0.379090	0.332115	0.149916	0.208133	0.043062
bmi	0.185085	0.088161	1.000000	0.395411	0.249777	0.261170	-0.366811	0.413807	0.446157	0.388680	0.586450
bp	0.335428	0.241010	0.395411	1.000000	0.242464	0.185548	-0.178762	0.257650	0.393480	0.390430	0.441482
s1	0.260061	0.035277	0.249777	0.242464	1.000000	0.896663	0.051519	0.542207	0.515503	0.325717	0.212022
s2	0.219243	0.142637	0.261170	0.185548	0.896663	1.000000	-0.196455	0.659817	0.318357	0.290600	0.174054
s3	-0.075181	-0.379090	-0.366811	-0.178762	0.051519	-0.196455	1.000000	-0.738493	-0.398577	-0.273697	-0.394789
s4	0.203841	0.332115	0.413807	0.257650	0.542207	0.659817	-0.738493	1.000000	0.617859	0.417212	0.430453
s5	0.270774	0.149916	0.446157	0.393480	0.515503	0.318357	-0.398577	0.617859	1.000000	0.464669	0.565883
s6	0.301731	0.208133	0.388680	0.390430	0.325717	0.290600	-0.273697	0.417212	0.464669	1.000000	0.382483
target	0.187889	0.043062	0.586450	0.441482	0.212022	0.174054	-0.394789	0.430453	0.565883	0.382483	1.000000

Feature Selection

- Use only the training dataset or the entire dataset for feature selection?

	age	sex	bmi	bp	s1	s2	s3	s4	s5	s6	target
age	1.000000	0.173737	0.185085	0.335428	0.260061	0.219243	-0.075181	0.203841	0.270774	0.301731	0.187889
sex	0.173737	1.000000	0.088161	0.241010	0.035277	0.142637	-0.379090	0.332115	0.149916	0.208133	0.043062
bmi	0.185085	0.088161	1.000000	0.395411	0.249777	0.261170	-0.366811	0.413807	0.446157	0.388680	0.586450
bp	0.335428	0.241010	0.395411	1.000000	0.242464	0.185548	-0.178762	0.257650	0.393480	0.390430	0.441482
s1	0.260061	0.035277	0.249777	0.242464	1.000000	0.896663	0.051519	0.542207	0.515503	0.325717	0.212022
s2	0.219243	0.142637	0.261170	0.185548	0.896663	1.000000	-0.196455	0.659817	0.318357	0.290600	0.174054
s3	-0.075181	-0.379090	-0.366811	-0.178762	0.051519	-0.196455	1.000000	-0.738493	-0.398577	-0.273697	-0.394789
s4	0.203841	0.332115	0.413807	0.257650	0.542207	0.659817	-0.738493	1.000000	0.617859	0.417212	0.430453
s5	0.270774	0.149916	0.446157	0.393480	0.515503	0.318357	-0.398577	0.617859	1.000000	0.464669	0.565883
s6	0.301731	0.208133	0.388680	0.390430	0.325717	0.290600	-0.273697	0.417212	0.464669	1.000000	0.382483
target	0.187889	0.043062	0.586450	0.441482	0.212022	0.174054	-0.394789	0.430453	0.565883	0.382483	1.000000

Calculate correlation matrix using the entire dataset

↓

...lect the 'BM...

↓

Split the ent... dataset into training ...st... dataset

↓

Mo... fit to training d...aset

↓

Model evaluation with testing dataset

Feature Selection

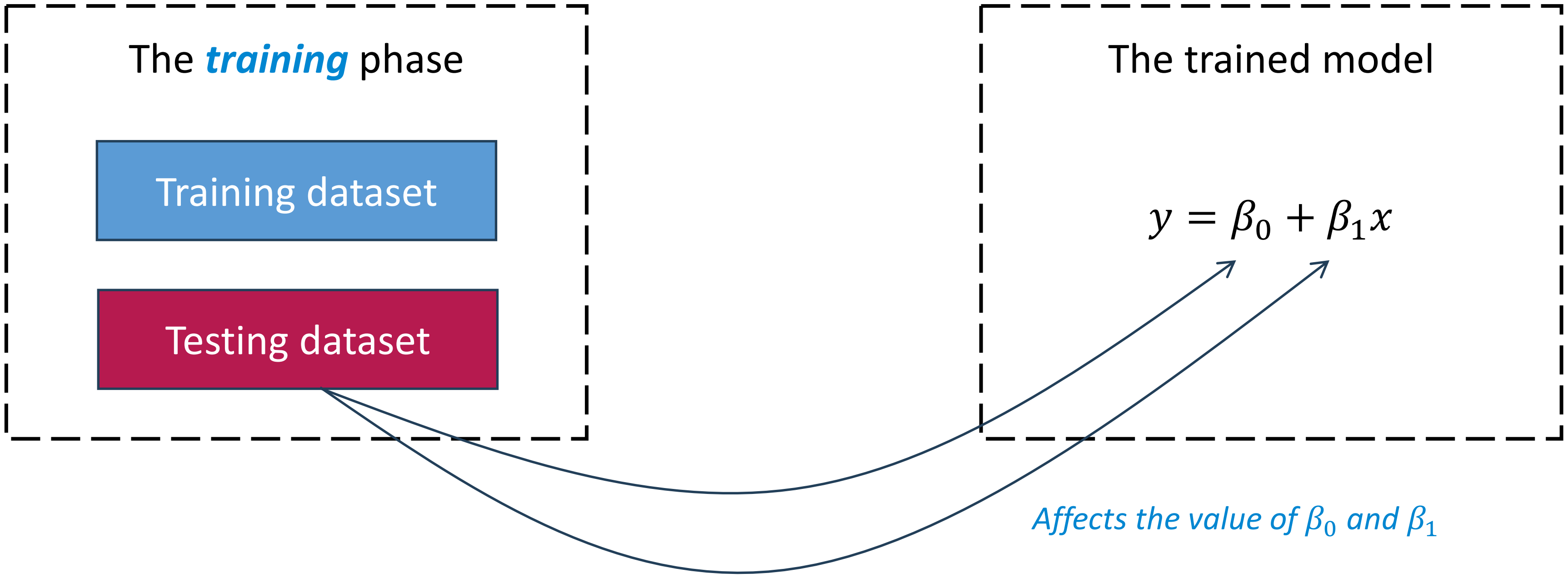
- Use only the training dataset for feature selection
- The goal of evaluation is to assess model performance on unseen data
- Make sure the model has never seen the testing dataset before evaluation

↓

- Avoid any impact of the testing dataset on the trained model

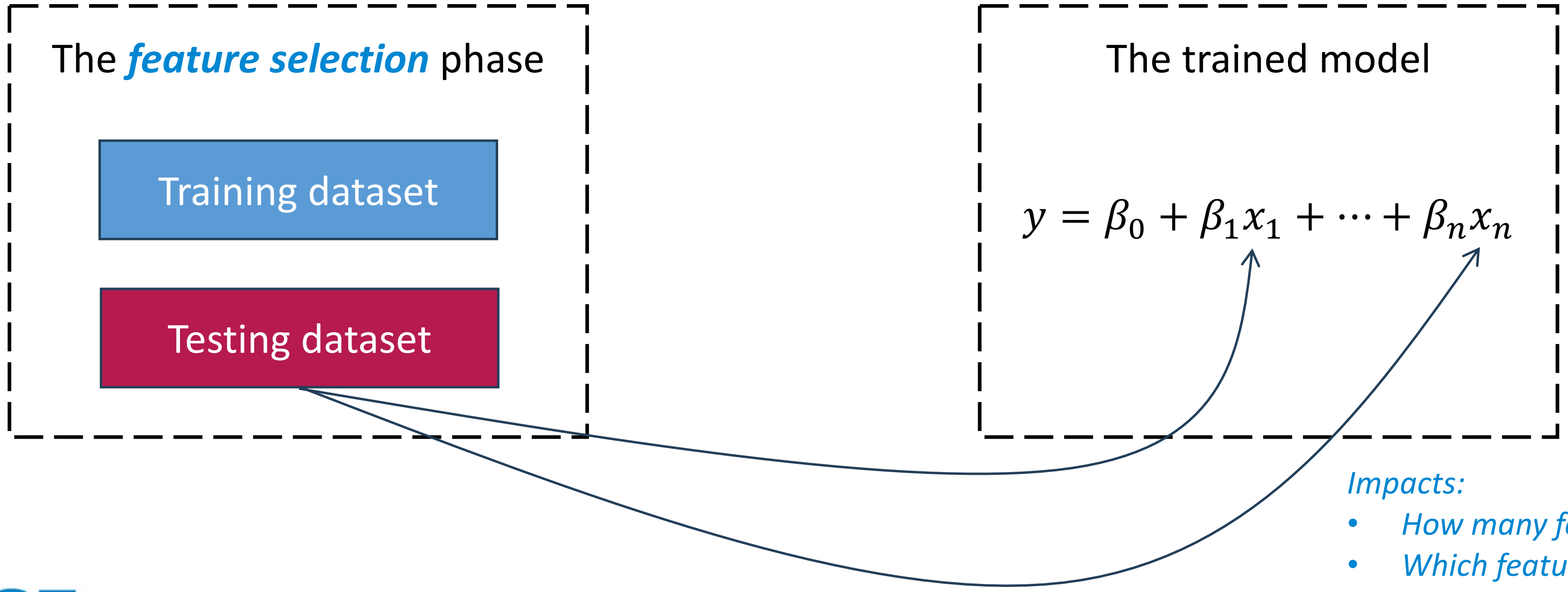
Feature Selection

- Use only the training dataset for feature selection
- Avoid any impact of the testing dataset on the trained model



Feature Selection

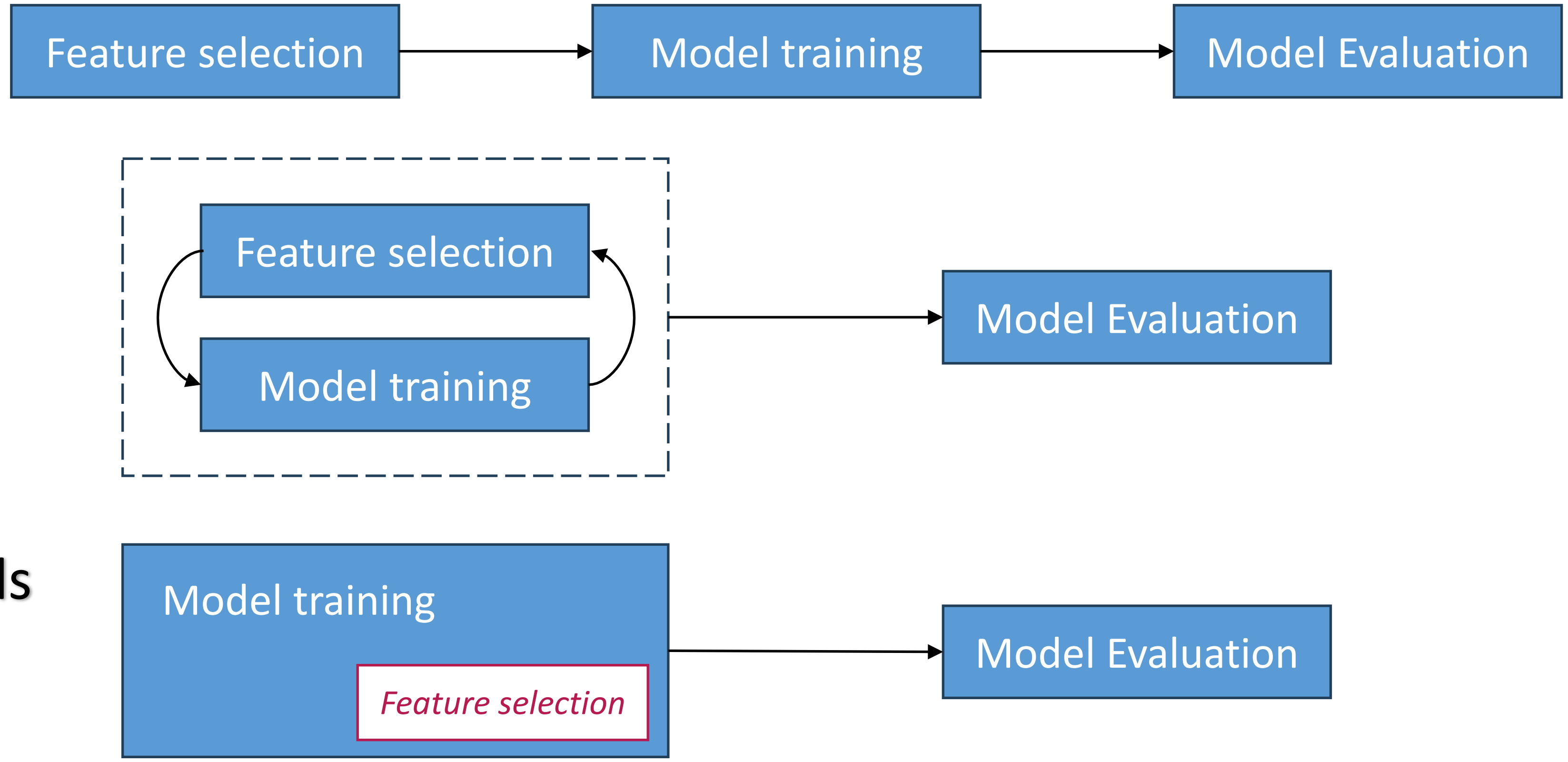
- Use only the training dataset for feature selection
- Avoid any impact of the testing dataset on the trained model



Impacts:
- *How many features to select*
- *Which feature to select*

Feature Selection

- Type of feature selection methods
  - Filter methods
  - Wrapper methods
  - Embedded methods



Feature Selection

- Type of feature selection methods
  - Filter methods

Feature selection → Model training → Model Evaluation

- Select features regardless of the model
- Only based on characteristics of the data (like correlation between feature and target)
- Need a score function to calculate a score for each feature

	MedInc	HouseAge	AveRooms	AveBedrms	Population	MedHouseVal
MedHouseVal	0.65	0.37	0.46	0.43	0.01	1.00

Select 3 features with the highest scores

Feature Selection

- Type of feature selection methods
  - Filter methods

Feature selection → Model training → Model Evaluation

- Tend to select redundant features
  - The score of each feature is determined only by the feature and the target variable
  - Didn't consider the relationships between different features

	MedInc	HouseAge	AveRooms	AveBedrms	Population	MedHouseVal
MedHouseVal	0.65	0.37	0.46	0.43	0.01	1.00

The correlation between these two features can be 0.95

KEDGE BUSINESS SCHOOL

Feature Selection

- Type of feature selection methods
  - Filter methods

Feature selection → Model training → Model Evaluation

- Didn't consider the specific algorithm for training the model
  - The result of feature selection is the same no matter what algorithm is used to train the model
  - Selecting features with high scores is not always beneficial for improving model performance

	MedInc	HouseAge	AveRooms	AveBedrms	Population	MedHouseVal
MedHouseVal	0.65	0.37	0.46	0.43	0.01	1.00

Feature with high score might be useless for some algorithms

Feature with low score might be useful for some algorithms

Feature Selection

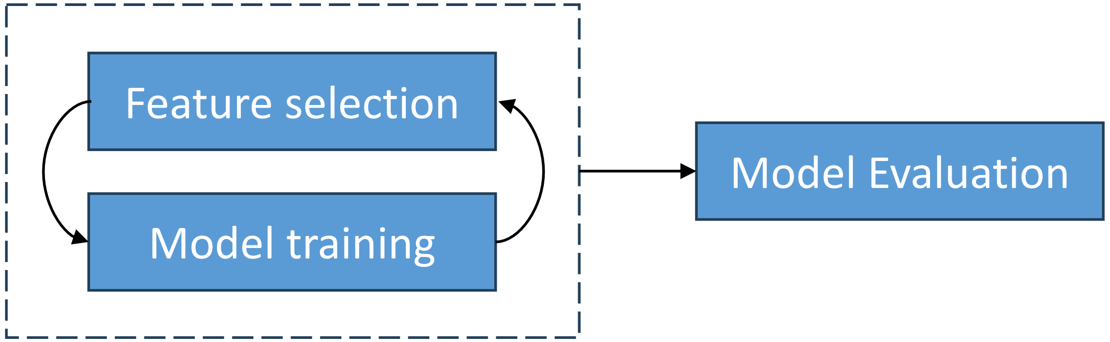
- Type of feature selection methods
  - Filter methods

    Feature selection → Model training → Model Evaluation

    - Score functions for regression model
      - `sklearn.feature_selection.r_regression`
        - Pearson correlation coefficient, $r \in [-1, 1]$
      - `sklearn.feature_selection.f_regression`
        - F – test statistic, $F \geq 0$, higher value is better
      - `sklearn.feature_selection.mutual_info_regression`
        - Mutual information, $MI \geq 0$, higher value is better

Feature Selection

- Type of feature selection methods
  - Wrapper methods



    - Evaluate different subsets of features by comparing model performance on training dataset

Subset 1: x_1 → *Training MSE: 2.1*

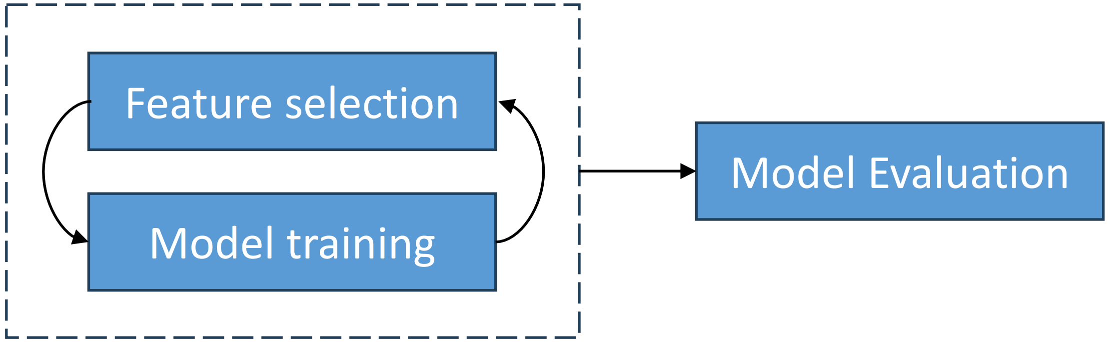
Subset 2: x_1, x_2 → *Training MSE: 1.3*

Subset 3: x_1, x_3 → *Training MSE: 0.7*

The best subset with the best model performance on the training dataset

Feature Selection

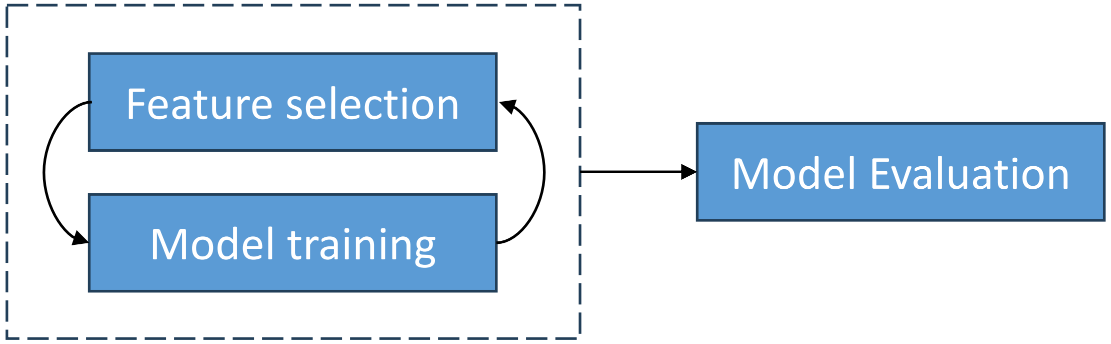
- Type of feature selection methods
  - Wrapper methods



- Evaluate different subsets of features by comparing model performance on training dataset
  - Enumeration
    - Enumerate all the possible subsets, select the best one

Feature Selection

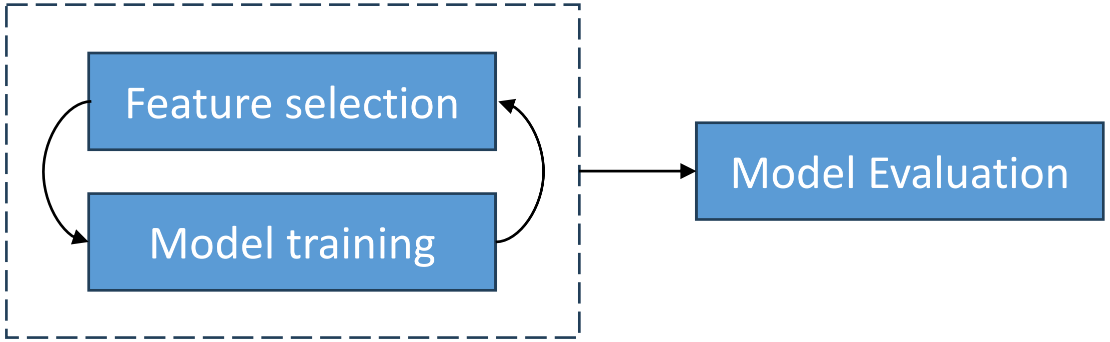
- Type of feature selection methods
  - Wrapper methods



- Evaluate different subsets of features by comparing model performance on training dataset
  - Forward selection - An iterative method
    - Start with having no feature
    - In each iteration, add the feature which best improves the model performance
    - Stop when
      - The pre-defined number of selected features is reached
      - An addition of a new feature cannot improve the model performance anymore

Feature Selection

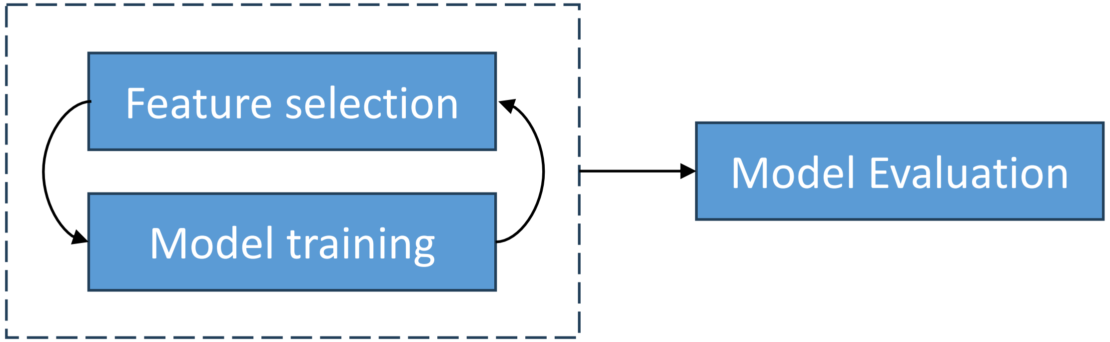
- Type of feature selection methods
  - Wrapper methods



- Evaluate different subsets of features by comparing model performance on training dataset
  - Backward elimination - An iterative method
    - Start with all features
    - In each iteration, eliminate the feature which best improves the model performance
    - Stop when
      - The pre-defined number of selected features is reached
      - An elimination of a feature cannot improve the model performance anymore

Feature Selection

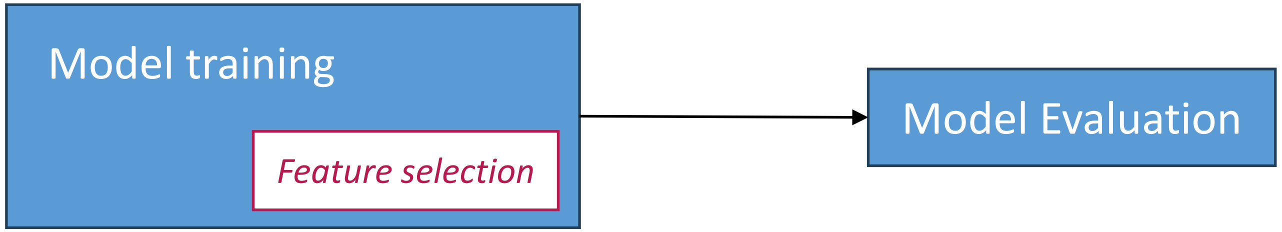
- Type of feature selection methods
  - Wrapper methods



- Advantage
  - Consider the interaction between different features
  - Consider the usefulness of the feature regarding to the specific algorithm
- Disadvantage
  - Increasing computation time significantly

Feature Selection

- Type of feature selection methods
  - Embedded methods



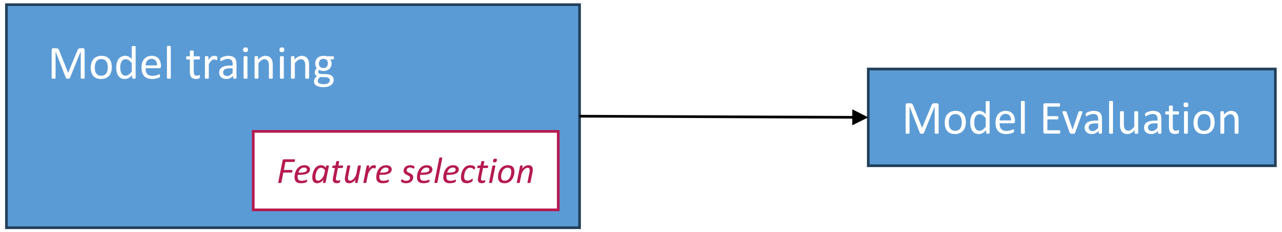
- Some algorithms embed the feature selection process into its learning process
  - Train the model on the training dataset with all features
  - The algorithm can automatically ignore the irrelevant features during the learning process
    - For example, [Lasso Regression] can assign close-to-zero coefficients to irrelevant features

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_N x_N$$

Irrelevant features

Feature Selection

- Type of feature selection methods
  - Embedded methods



- Some algorithms embed the feature selection process into its learning process
  - Train the model on the training dataset with all features
  - The algorithm can automatically ignore the irrelevant features during the learning process
    - For example, [Lasso Regression]() can assign close-to-zero coefficients to irrelevant features

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_N x_N$$

Value close to zero

Outline

- Multiple Linear Regression
- Feature Selection
- **Polynomial Regression**
- Under-fitting & Over-fitting

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

$$\beta_0 \quad + \quad \beta_1 \quad \times \quad x \quad - \quad \beta_2 \quad \times \quad x^2$$

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

$$\beta_0 \quad + \quad \beta_1 \quad \times \quad x \quad - \quad \beta_2 \quad \times \quad x^2$$

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

$$\beta_0 \quad + \quad \beta_1 \quad \times \quad x \quad - \quad \beta_2 \quad \times \quad x^2$$

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

$$\beta_0 \quad + \quad \beta_1 \quad \times \quad x \quad - \quad \beta_2 \quad \times \quad x^2$$

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

$$\beta_0 \quad + \quad \beta_1 \quad \times \quad x \quad - \quad \beta_2 \quad \times \quad x^2$$

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

$$\beta_0 \quad + \quad \beta_1 \quad \times \quad x \quad - \quad \beta_2 \quad \times \quad x^2$$

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

$$x^0 \times \beta_0 + \beta_1 \times x - \beta_2 \times x^2$$

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

A term

$$x^0 \times \beta_0 + \beta_1 \times x - \beta_2 \times x^2$$

  - Polynomials appear as a sequence of terms, where each term has a coefficient and a variable raised to a power, which is a non-negative integer

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

A sequence of three terms

$$x^0 \times \beta_0 + \beta_1 \times x - \beta_2 \times x^2$$

  - Polynomials appear as a sequence of terms, where each term has a coefficient and a variable raised to a power, which is a non-negative integer

Polynomial Regression

- Polynomial
  - An expression consisting of coefficients and variables, that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponentiation of variables

The addition and subtraction operation exist between different terms

$$x^0 \times \beta_0 + \beta_1 \times x - \beta_2 \times x^2$$

  - Polynomials appear as a sequence of terms, where each term has a coefficient and a variable raised to a power, which is a non-negative integer

Polynomial Regression

- Polynomial
  - The degree of a polynomial is the highest power of the variable in the expression

Expression	$f(x) = 3$	$f(x) = 3 + x$	$f(x) = 3 - x^2$	$f(x) = x^3 + x^5$
Degree	0	1	2	5

Polynomial Regression

- Polynomial
  - A single variable in the polynomial expression
    - $f(x) = 3 - x + 2x^2$
  - There can be more than one variable in the polynomial expression
    - $f(x_1, x_2) = 3 - x_1 + 2x_2 + x_1x_2 - 2x_1^2 + 5x_2^2$

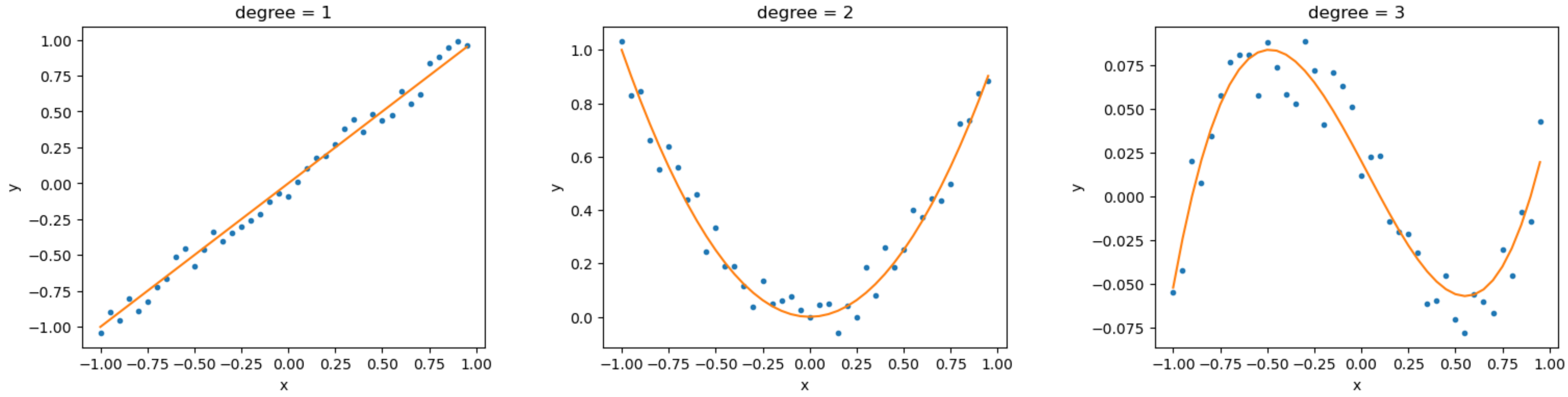
  - What is the degree of the following polynomial expression?
    - $f(x_1, x_2) = 3 - x_1x_2 + 5x_2^2 - 2x_1x_2^2$

Polynomial Regression

- Polynomial regression
  - Estimate the relationship between dependent variable and the independent variables as a polynomial expression
  - For the situation with a single independent variable
    - $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$
    - n is the degree of the polynomial regression

Polynomial Regression

- Polynomial regression extends linear regression with polynomial terms to model non-linear relationship



Polynomial Regression

- Polynomial regression can be seen as an extension of linear regression

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$$

Polynomial features

$$\downarrow$$

consider as

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_n z_3$$

New features

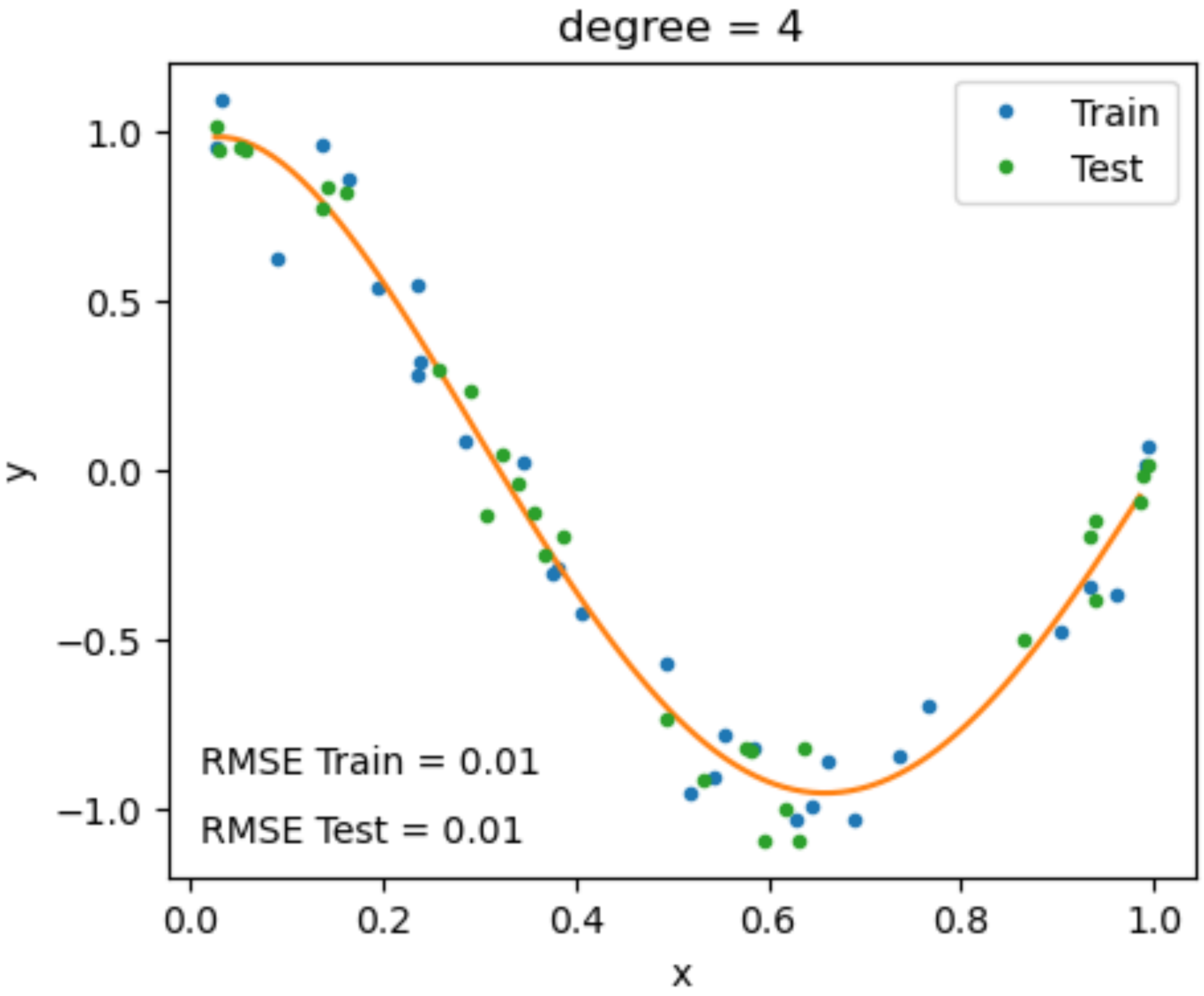
- We can use multiple linear regression to implement polynomial regression by adding polynomial features as new features

Outline

- Multiple Linear Regression
- Feature Selection
- Polynomial Regression
- **Under-fitting & Over-fitting**

Under-fitting & Over-fitting

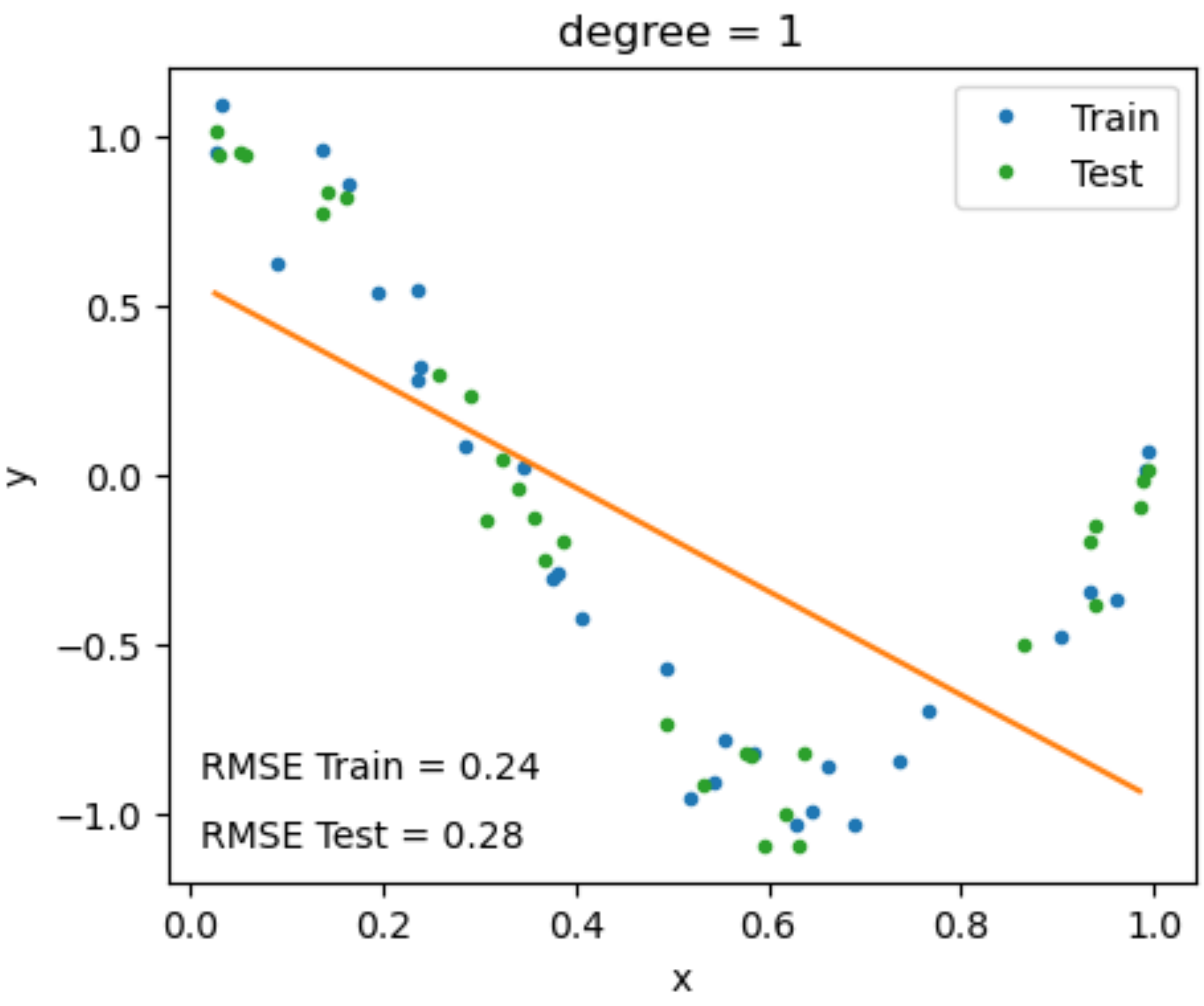
- What is a well-fitted model?
  - The model fits well on the training dataset
    - Low prediction error with training dataset
  - The model generalizes well on the testing dataset
    - Low prediction error with testing dataset



KEDGE BUSINESS SCHOOL

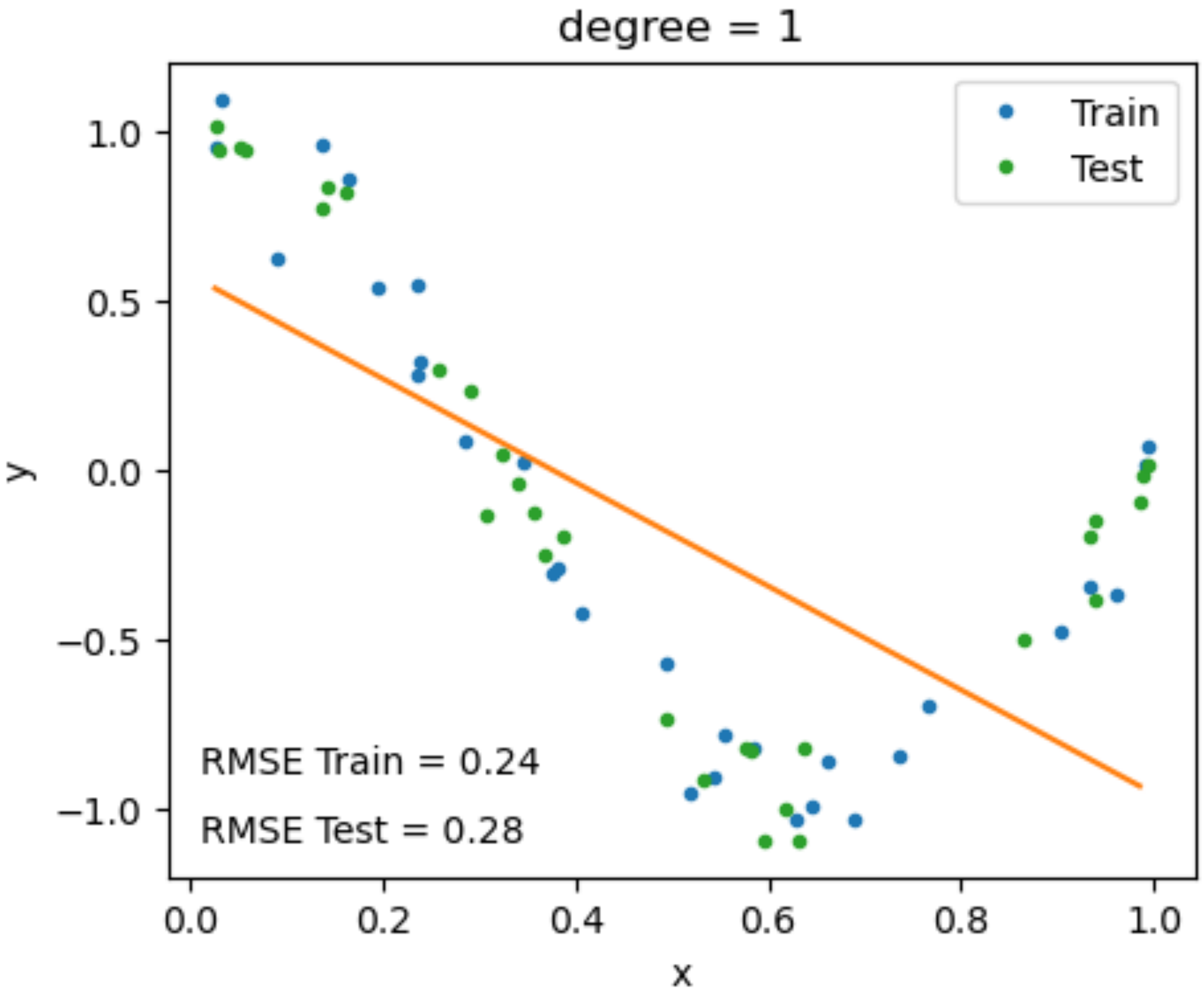
Under-fitting & Over-fitting

- Under-fitting
  - The model didn't fit well on the training dataset
    - High prediction error with training dataset
  - The model can't generalize well on the testing dataset
    - High prediction error with testing dataset



Under-fitting & Over-fitting

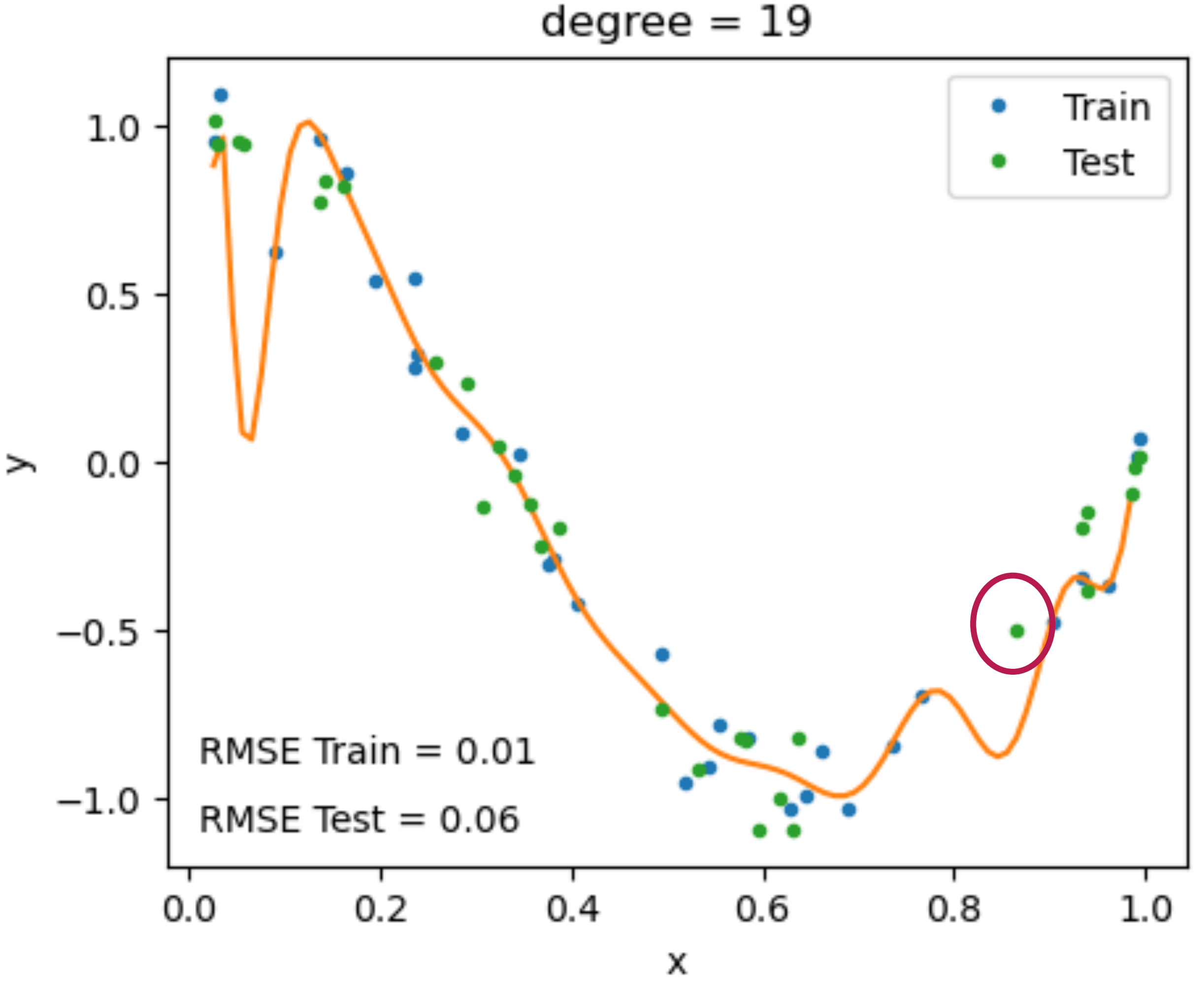
- Under-fitting
  - An under-fitted model is too simple compared to the complexity of the relationship to fit.
  - It fails to capture the underlying patterns, resulting in poor performance on both training and testing dataset.
  - It can't capture the training data trends, let alone generalize to new data.



KEDGE BUSINESS SCHOOL

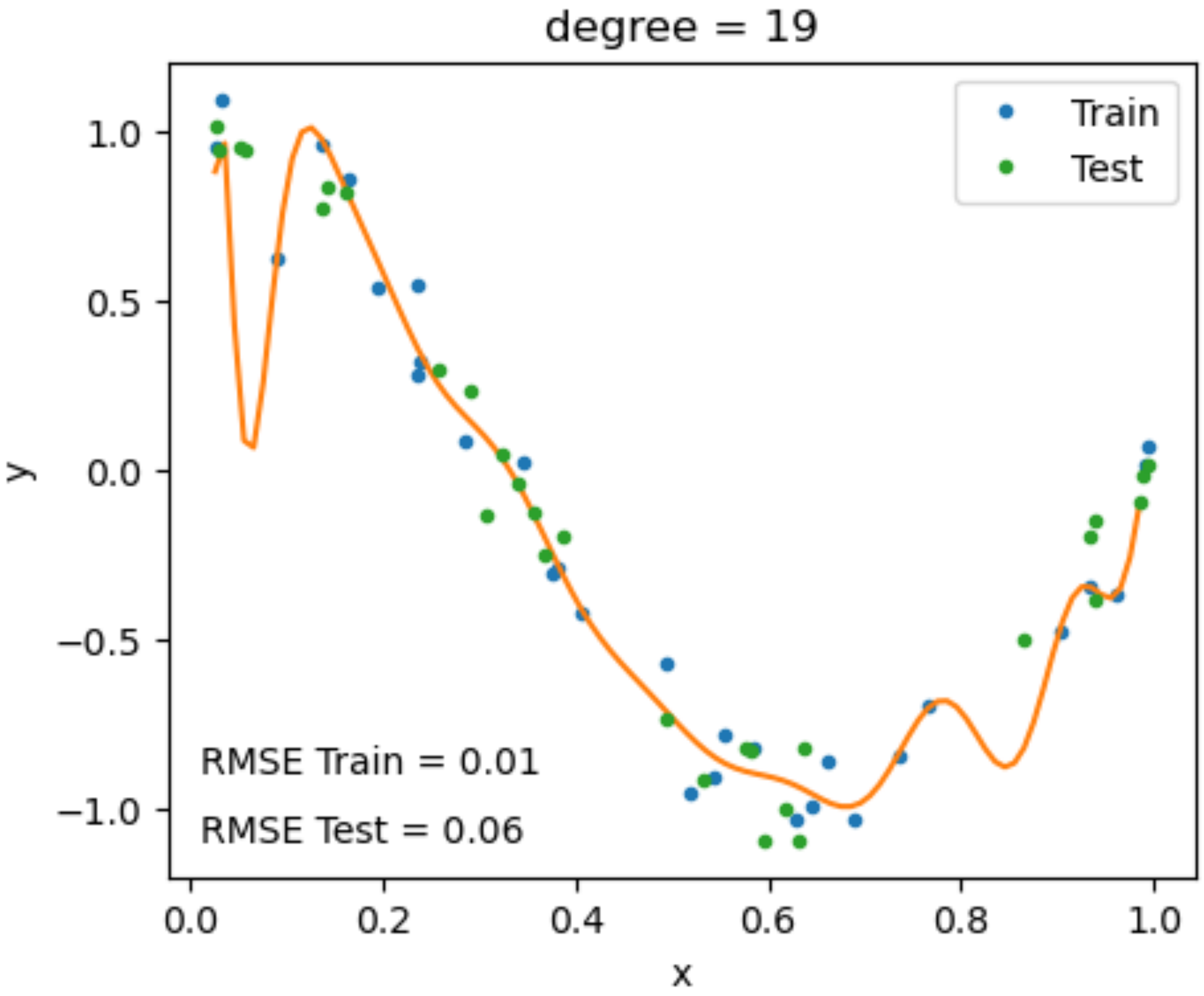
Under-fitting & Over-fitting

- Over-fitting
  - The model fitted too well on the training dataset
    - Low prediction error with training dataset
  - The model didn't generalize well on the testing dataset
    - High prediction error with testing dataset



Under-fitting & Over-fitting

- Over-fitting
  - The over-fitted model is too complex, with too many parameters tailed to fit the fluctuations in the training dataset
  - It captures the noise in the training dataset along with the underlying pattern.

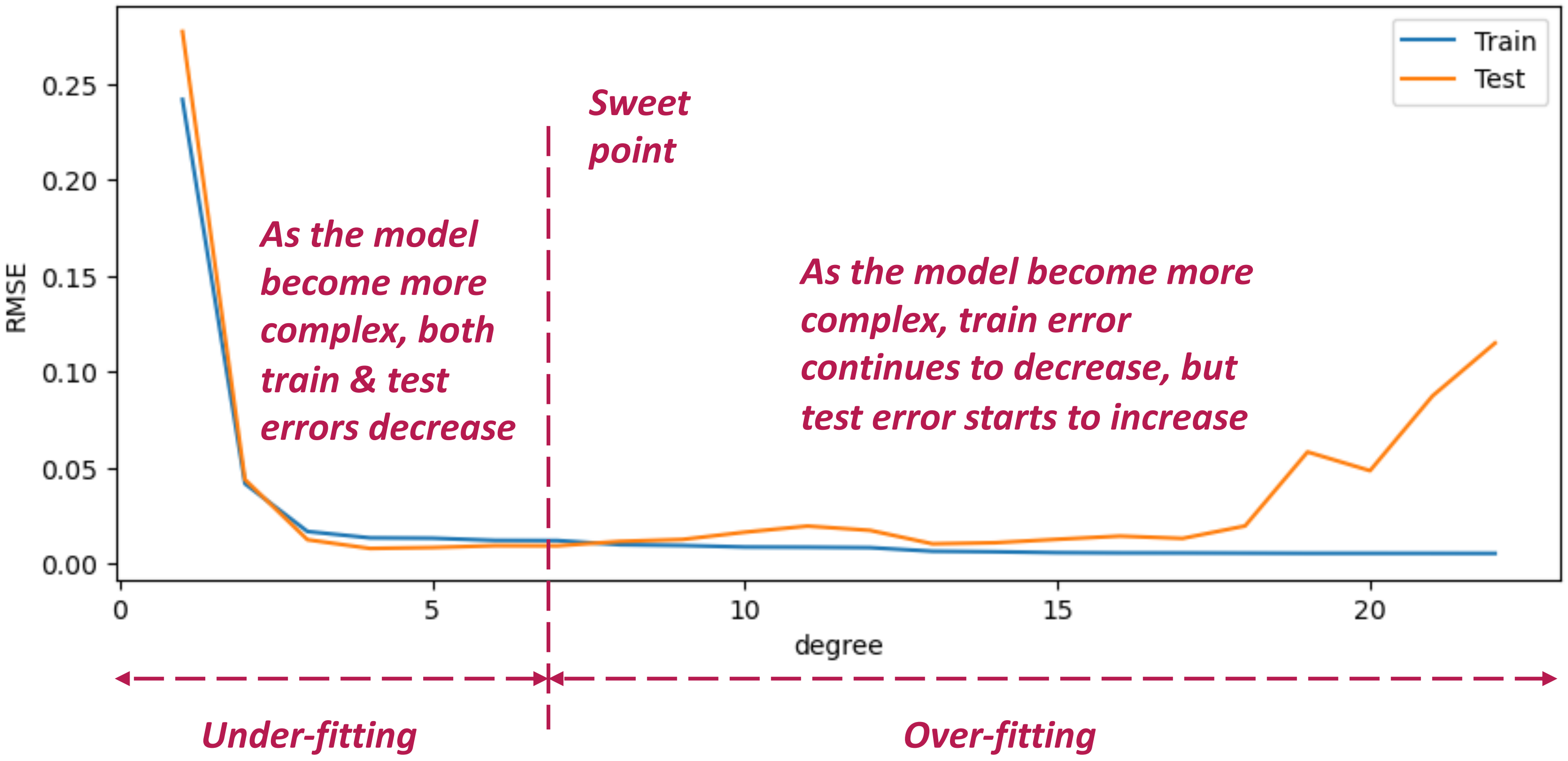


Under-fitting & Over-fitting

- Over-fitting
  - Imagine an infinitely complex model like a dictionary
    - You can find every piece of data in the training data set in this dictionary.
      - 100% accurate prediction on training data
    - But for new data that has not been seen before, there is no record in the dictionary.
      - Cannot make prediction on unseen data

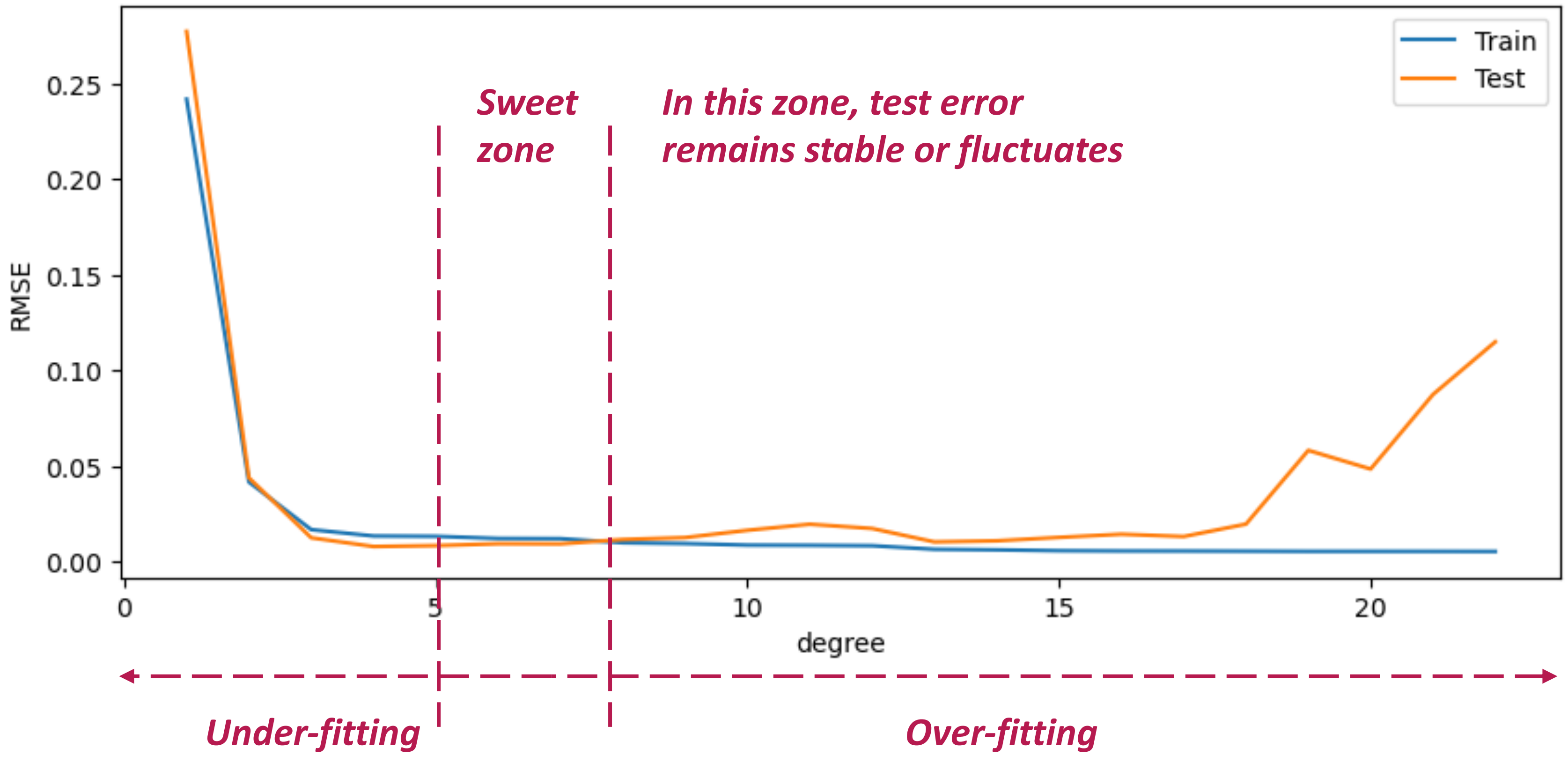
Under-fitting & Over-fitting

- Compare train & test errors to identify under- & over-fitting



Under-fitting & Over-fitting

- Compare train & test errors to identify under- & over-fitting



Hands-on Exercises

- Exercise 03 Regression II

KEDGE BUSINESS SCHOOL