Overfitting and Generalization

Best Model Selection

- **We can build various models**
  - K-NN: We may choose various values for k
  - Additive Linear Model: We may user polynomials of various order

- **But.. I want to build the best model**

Best Model Selection

- **Various models for one input**

$$f(x) = w_0$$

$$f(x) = w_1 x + w_0$$

$$f(x) = w_2 x^2 + w_1 x + w_0$$

$$f(x) = w_3 x^3 + w_2 x^2 + w_1 x + w_0$$

…

Model Complexity increases

Best Model Selection

- **Various Models for two inputs**

$$f(x_1, x_2) = w_0$$

$$f(x_1, x_2) = w_2 x_2 + w_1 x_1 + w_0$$

$$f(x_1, x_2) = w_5 x_2^2 + w_4 x_1^2 + w_3 x_1 x_2 + w_2 x_2 + w_1 x_1 + w_0$$

$$\begin{aligned} f(x_1, x_2) = {} & w_9 x_2^3 + w_8 x_1^3 + w_7 x_2^2 x_1 + w_6 x_2 x_1^2 \\ & + w_5 x_2^2 + w_4 x_1^2 + w_3 x_1 x_2 + w_2 x_2 + w_1 x_1 + w_0 \end{aligned}$$

…

Model Complexity increases

증가한다!

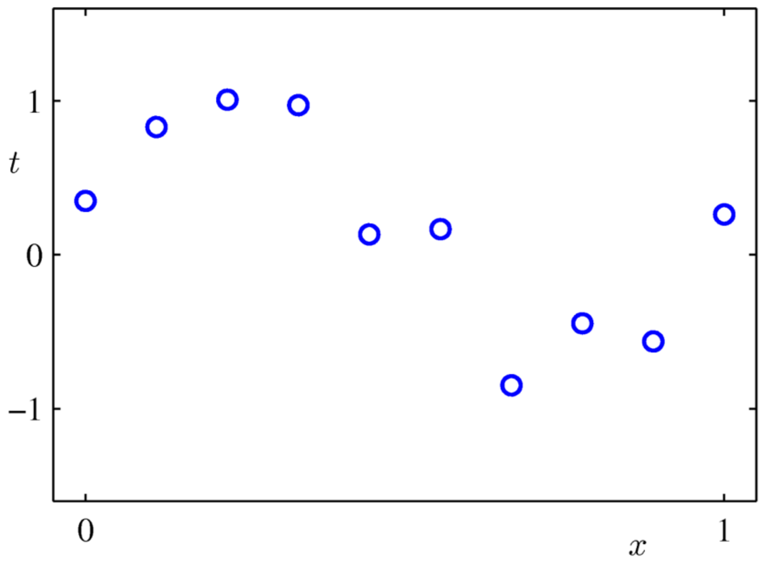
Best Model Selection

- **Example: Which model will be best for the data?**

  best의 의미?

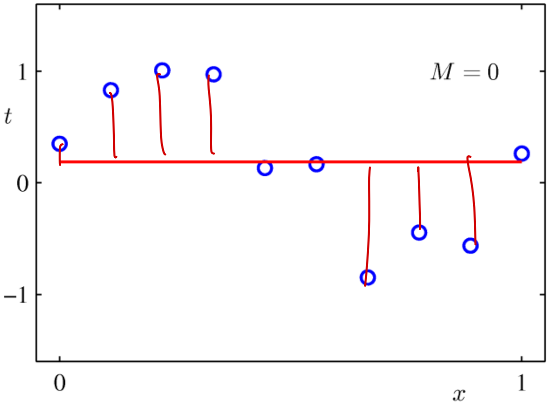
  - Zeroth order polynomial? First order polynomial?, or ..
  - Hmm.. Why don't we try all the possible models

Training set

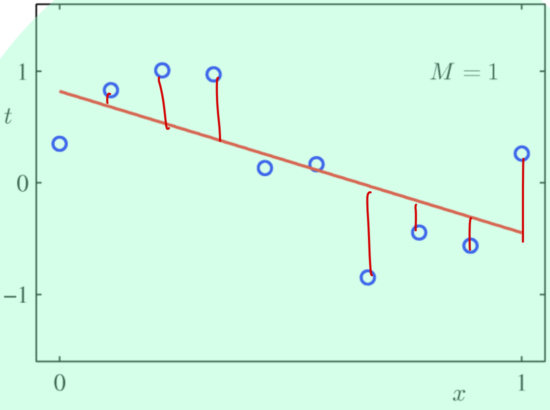


Overfitting vs Generalization

- **Which model will be best for the data?**
  - The model which has the least error as much as possible



0th order polynomial regression



1st order polynomial regression

This is better because it has less error

Training Data

past dataset

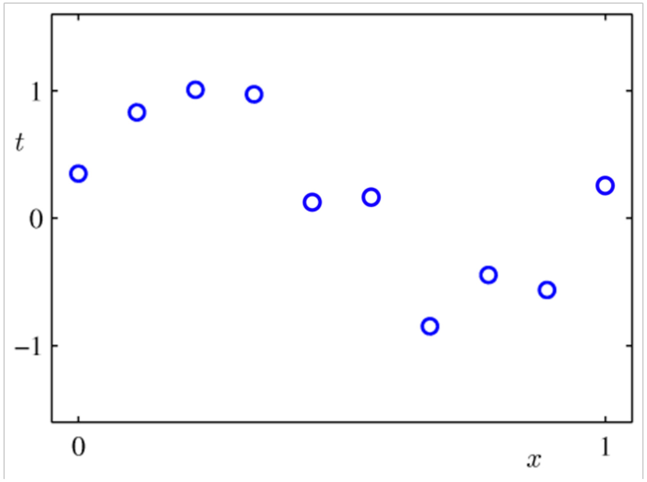
Build Model

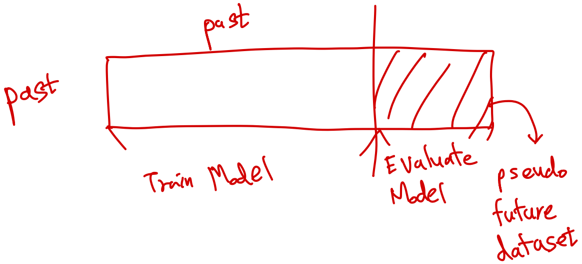
future dataset (unknown) → Test Data

Evaluate model

Overfitting vs Generalization

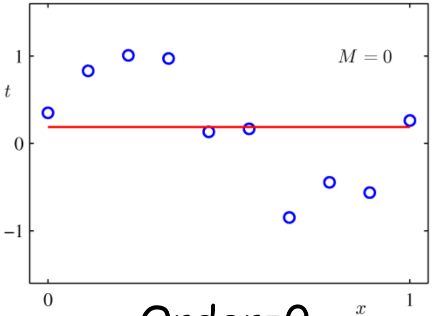
- **Which model will be best for the data?**





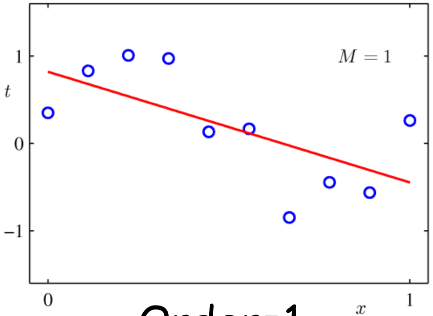
Training Error

Very Large



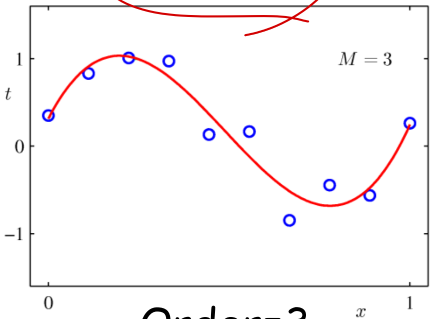
Order=0

Large



Order=1

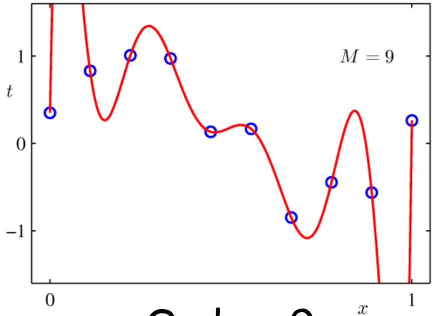
Modest



Order=3

Zero

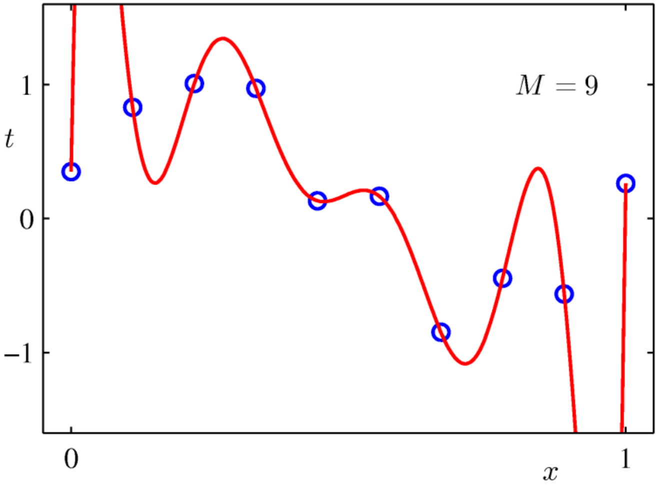
No Error!



Order=9

Overfitting vs Generalization

- **Which model will be best for the data?**
  - What about this?



9th order polynomial regression

- This may be the BEST because the training error is ZERO!!

Do you agree with this?

Overfitting vs Generalization

- **What is the purpose of Machine Learning?**

choice① X

Learning the given data as exactly as possible

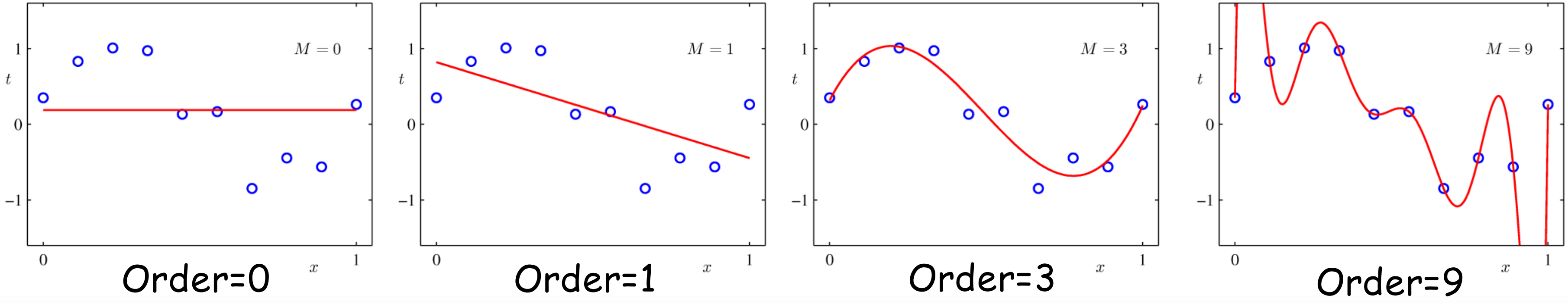
vs

choice② ✓

Predict the unknown data as exactly as possible based on the given data

= ML의 목적

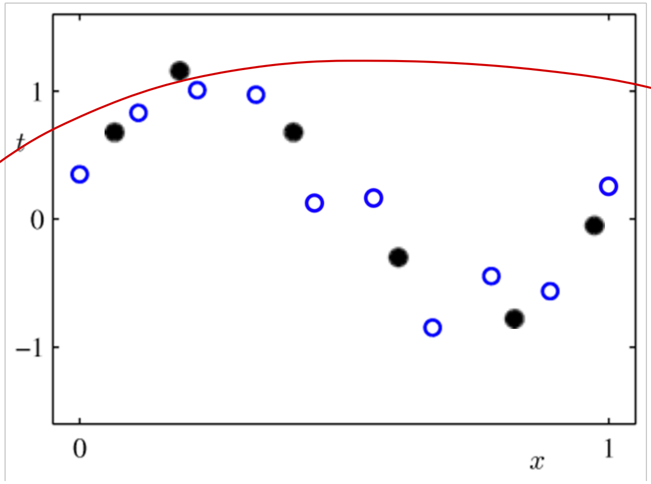
② ①



Order=0 Order=1 Order=3 Order=9

Overfitting vs Generalization

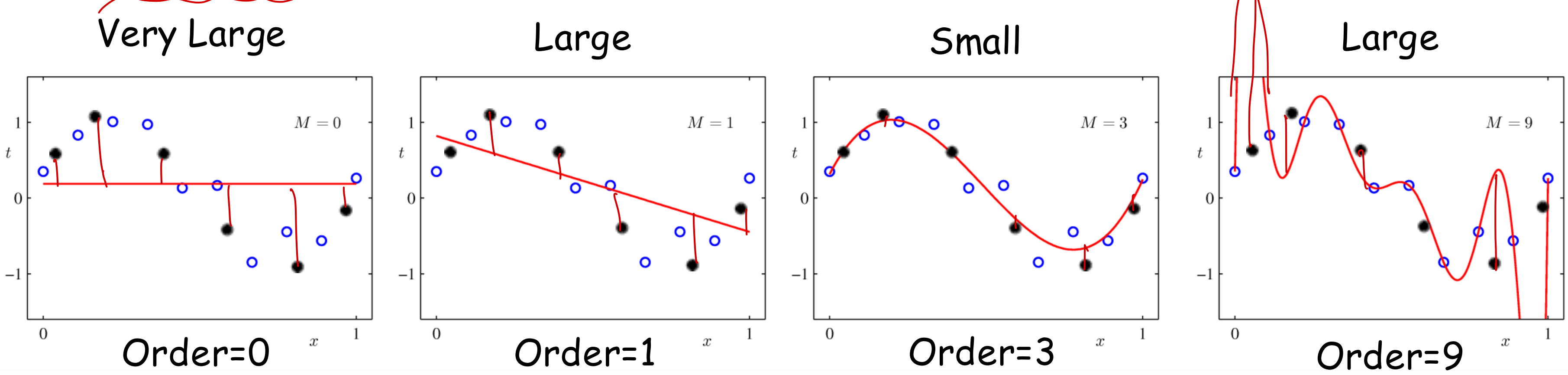
- **As the complexity of model increases,**



New samples

☆ 중요

Test Error



Very Large

Large

Small

Large

Order=0

Order=1

Order=3

Order=9

Overfitting vs Generalization

=best model

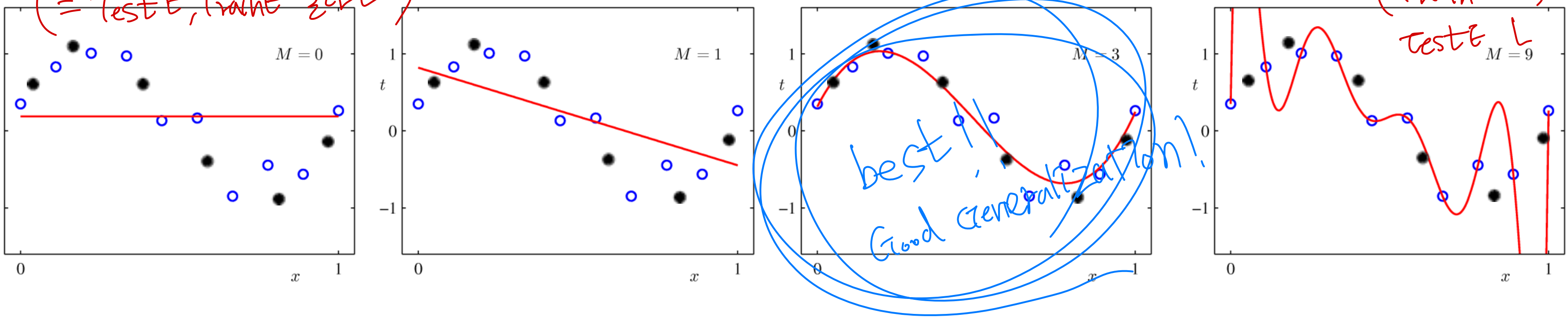
- **As the complexity of model increases,**
  - The model can more exactly learn the given data
  - However, the prediction accuracy (for the test set) does not necessarily increase

Test Error Large L S L 과적합

Underfit ↔ Overfit

(= Test E, Train E 둘다 L)

= (Train E S, Test E L)

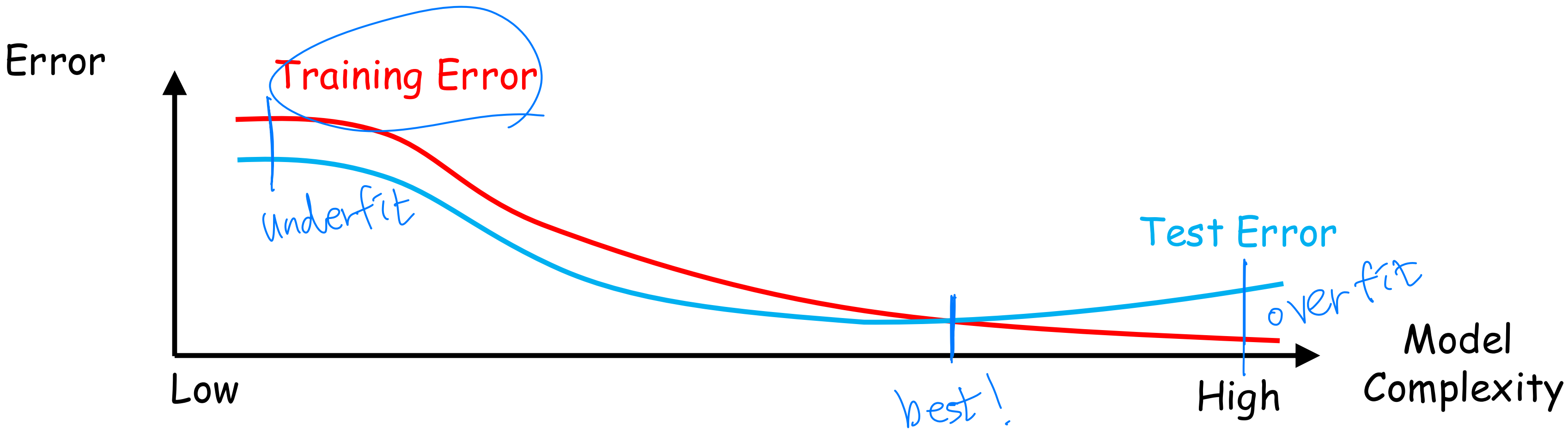


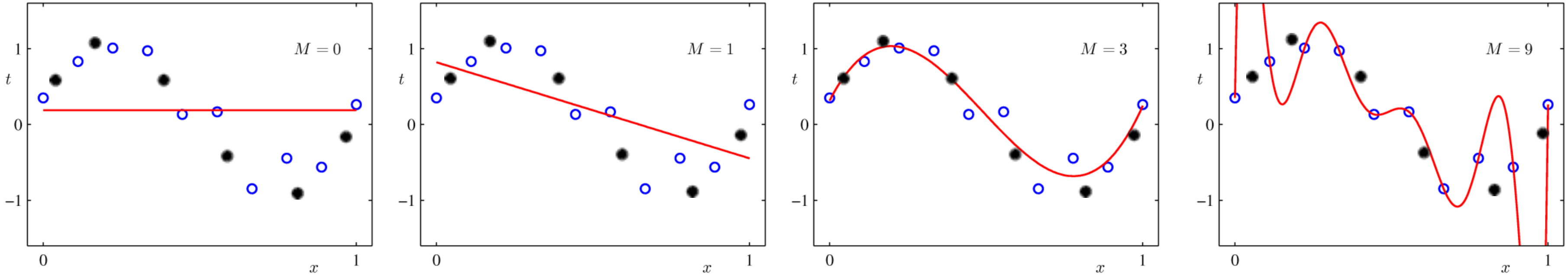
Low Complexity ↔ High Complexity

Training Error Large Small

Overfitting vs Generalization

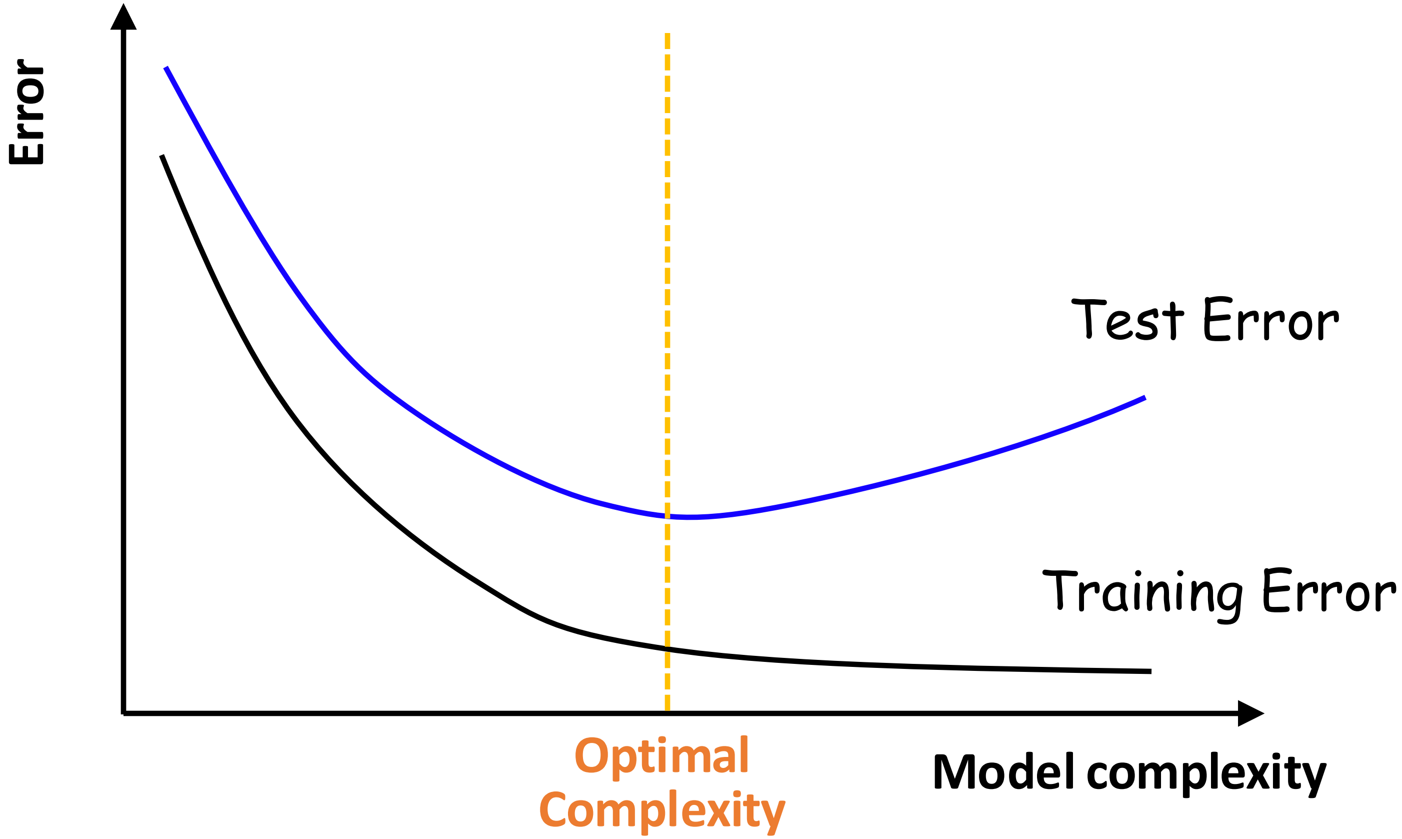
- **Overfitting**





Overfitting vs Generalization

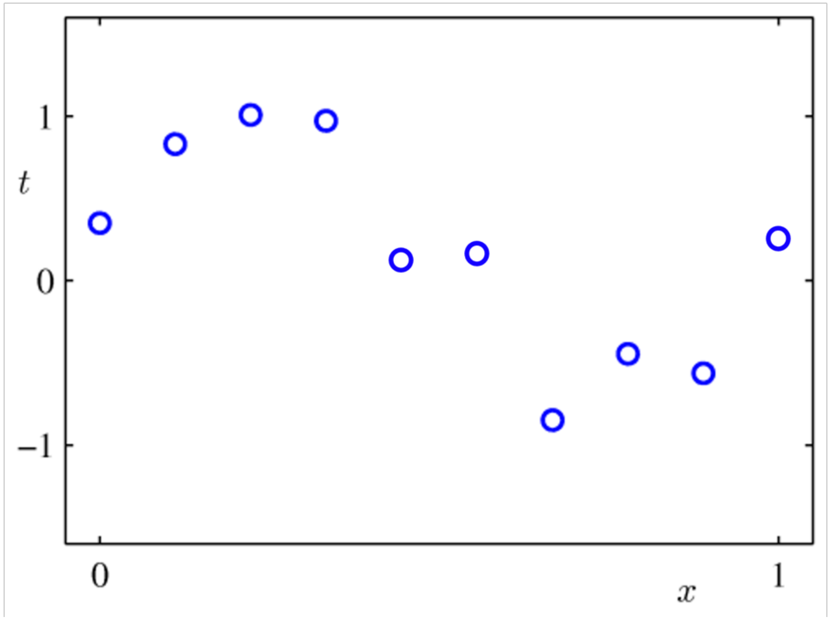
- **Hmm.. How can I choose the optimal?**

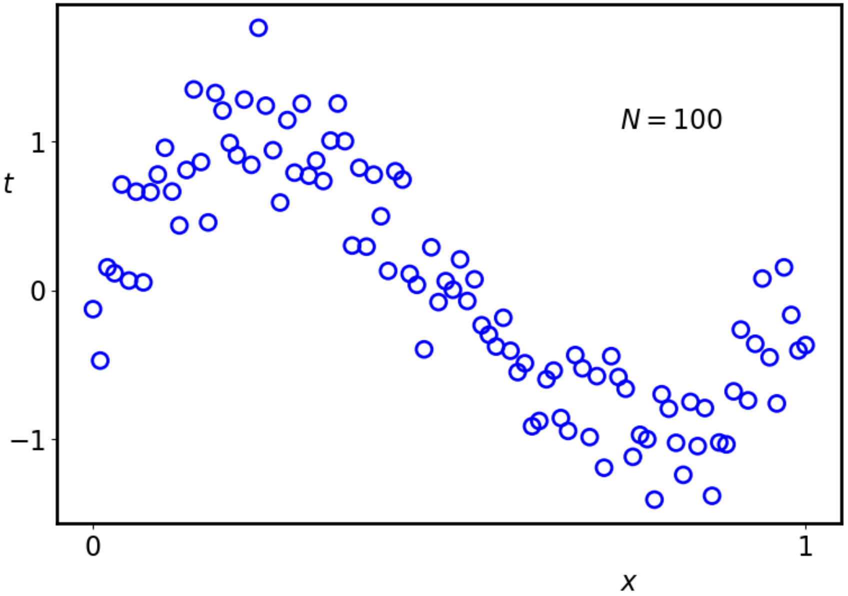


How to Achieve Good Generalization?

- **One of easiest way is.. Collecting More Data**

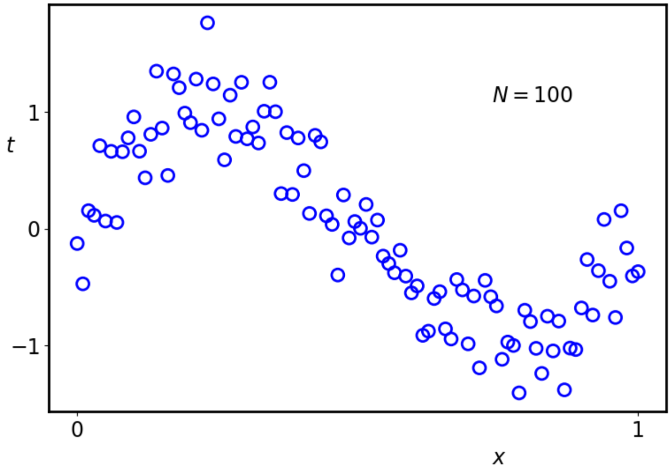
많은 데이터 모아라!





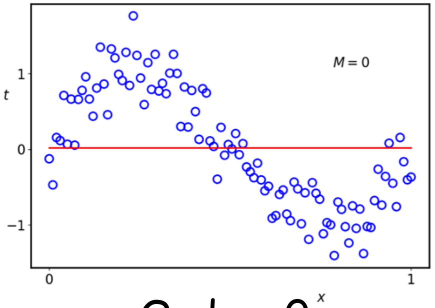
How to Achieve Good Generalization?

- **Model building with a large dataset**



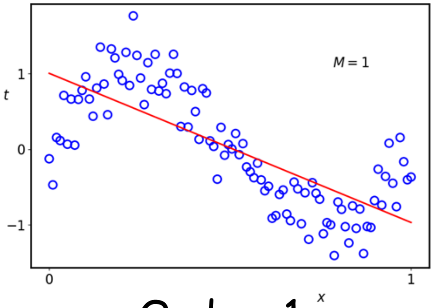
Training Error

Very Large



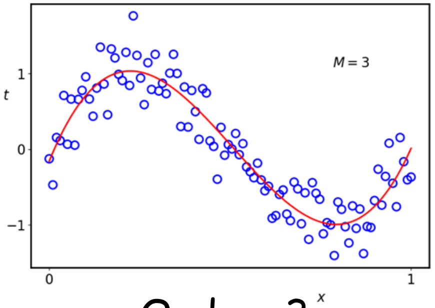
Order=0

Large



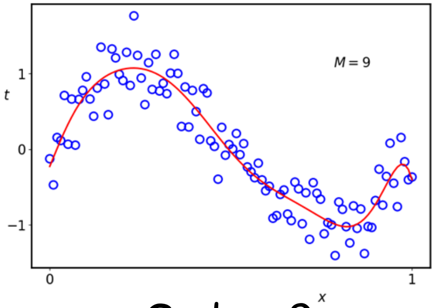
Order=1

Modest



Order=3

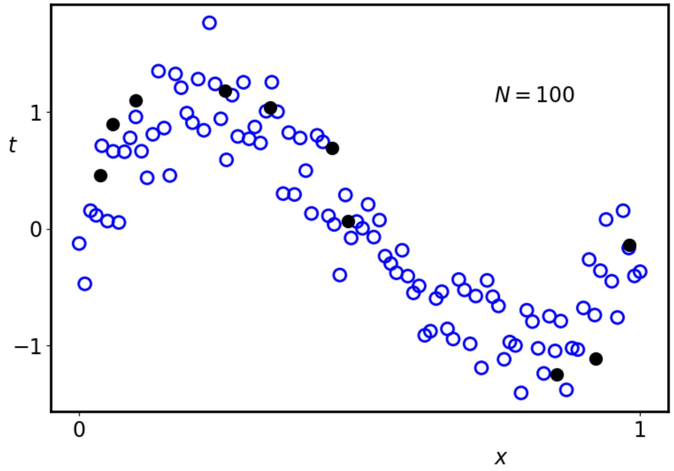
Modest



Order=9

How to Achieve Good Generalization?

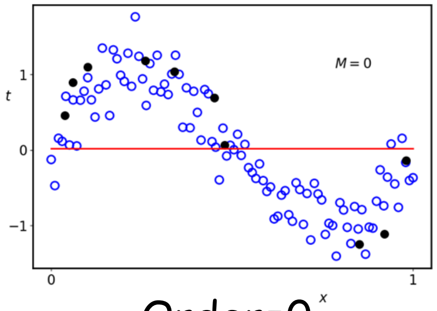
- **Model building with a large dataset**



Model is not overfitted!!

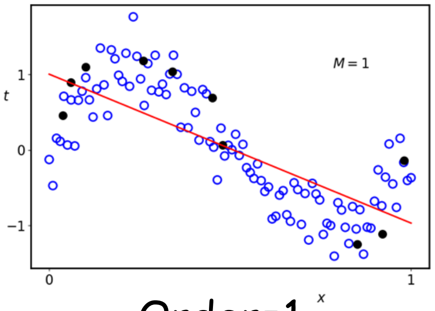
Test Error

Very Large



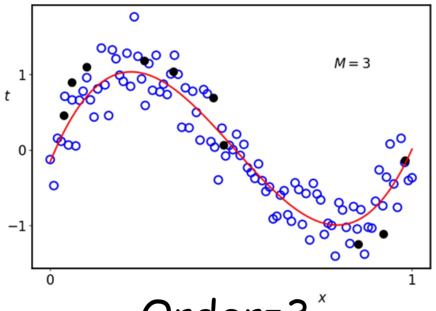
Order=0

Large



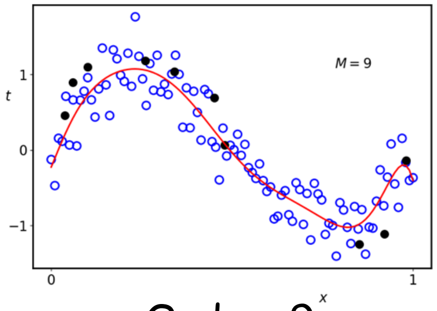
Order=1

Modest



Order=3

Modest



Order=9

How to Achieve Good Generalization?

- **How many samples to avoid overfitting?** example
  - Let’s assume that we need $\overset{10^2}{100}$ samples per an input feature (x, y)
  - If we have two features, we need 10^4 ← (x_1, x_2, y)
  - If we have three features, we need 10^6 ← (x_1, x_2, x_3, y)
  - If we have four features, we need 10^8
  - If we have ten features, we need 10^{20}

demension 높으면

너가 얼만큼의 데이터를 가졌는지 상관없이 overfit 피하기 힘들다

충분한

- **Though you have a substantial amount of data, you are uncertain whether the model is overfitted or not**

  불확실하다
  - Still, it may not be sufficient...

How to Achieve Good Generalization?

- **Build many models and choose the best**
  - Train many models
  - Evaluate them with Cross-Validation or Hold-out method
  - Choose the best

Build & selection

ch.5

- **Use Regularization Method**
  - There are many regularization methods
  - You may train a model with the regularization method

ch.6