國立清華大學
National Tsing Hua University

INNOVATION ENGINEERING LAB

智慧化企業整合
Intelligent Integration of Enterprise

Linear Regression

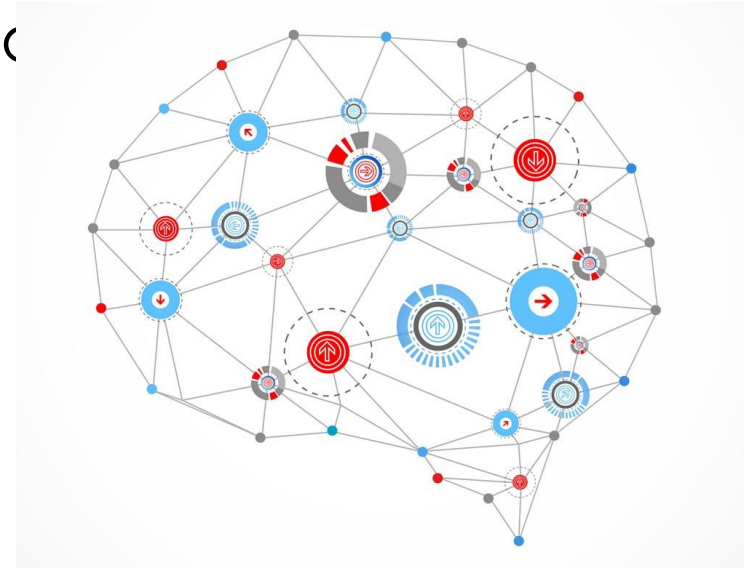
助教:陳可馨

Steps of Define Functions

Step1: Define your network (function) structure/ model

Step2: Measure the goodness o

Step3: Find the best function



國立清華大學
National Tsing Hua University

INNOVATION ENGINEERING LAB

Step1: Function Structure/ Model

國立清華大學 National Tsing Hua University

INNOVATION ENGINEERING LAB

Step 1: Model

ex: House Price Forecast

$y = b + w \cdot x$

w and b are parameters (can be any value)

A set of function

Model

$f_1, f_2 \cdots$

$f($  $) =$ Predicted House price

Linear model:

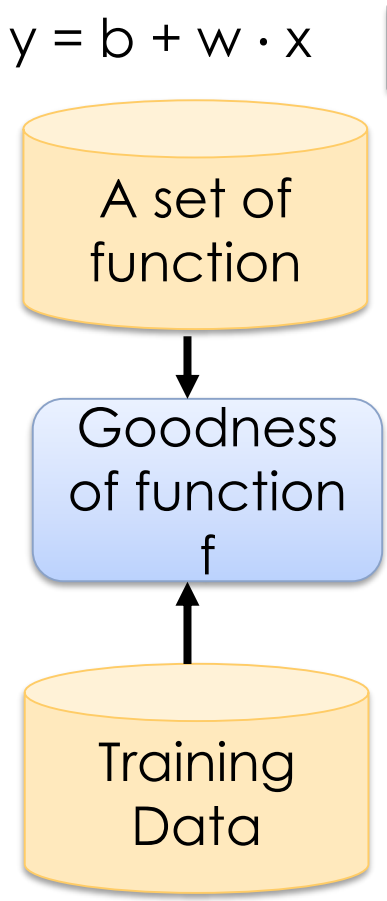
$$y = b + \sum w_i x_i$$

x_i: feature

w_i: weight, b: bias

Step2: Goodness of Function

Step 2: Goodness of Function



Loss function L:
Input: a function
Output: how bad it is

Estimation error

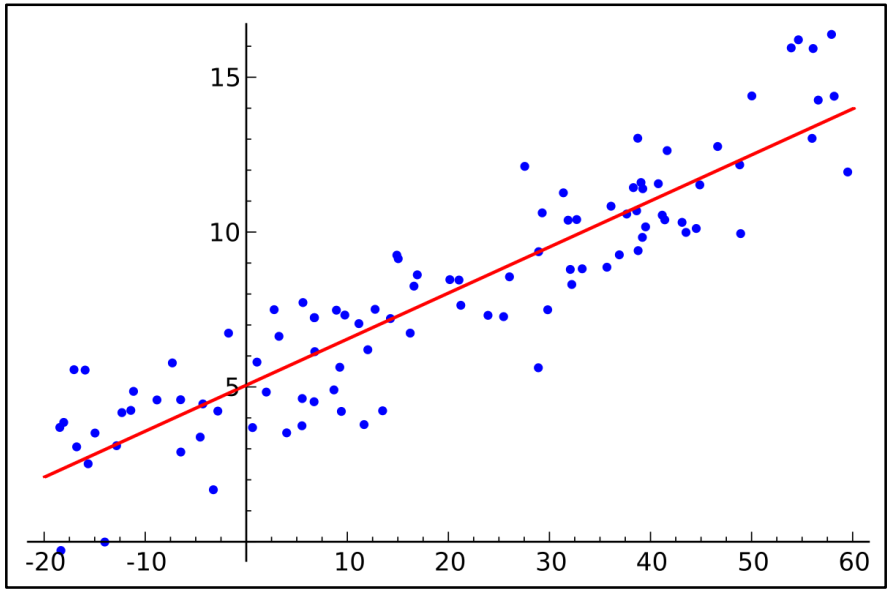
$$L(f) = \sum_{n=1}^{k} \left(\hat{y}^n - f(x^n)\right)^2$$

Sum over examples

Estimated y based on input function

$$L(w,b) = \sum_{n=1}^{k} \left(\hat{y}^n - (b + w \cdot x^n)\right)^2$$

Loss Function



- A measure of goodness
- Distance (error rate) between predictions and true labels
- Common loss functions: MSE, MAE, Cross entropy…

國立清華大學
National Tsing Hua University

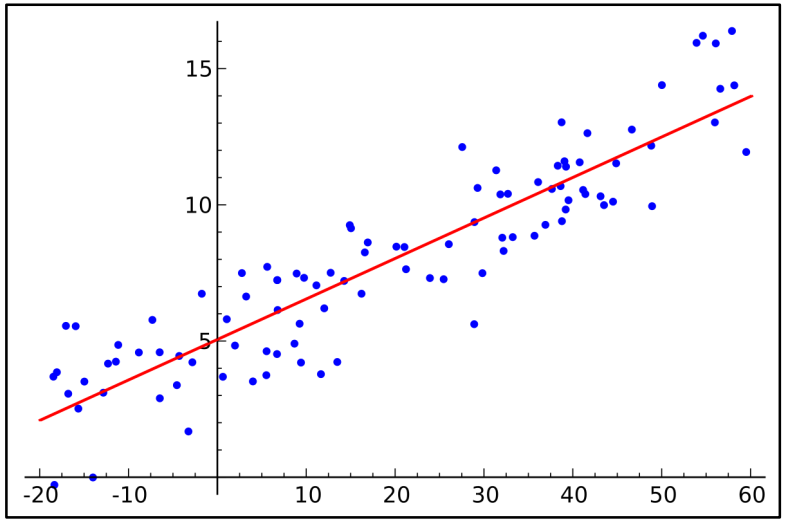
INNOVATION ENGINEERING LAB

Step3: The best Function

國立清華大學 National Tsing Hua University

INNOVATION ENGINEERING LAB

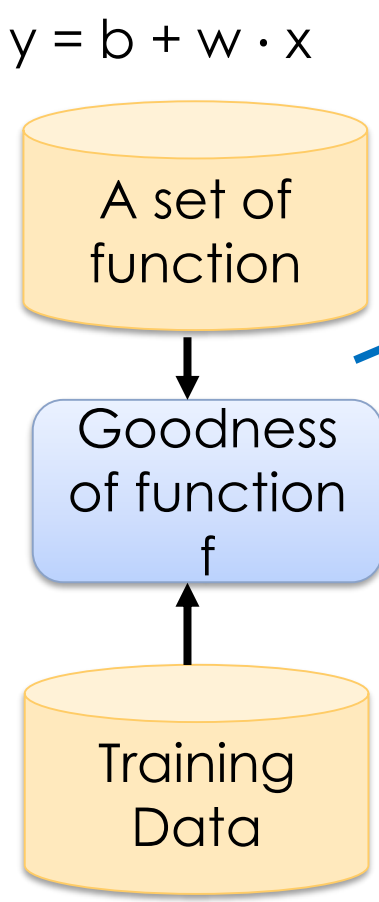
Goal



- Minimize loss function (EX: MSE)

$$L(w,b) = \sum_{n=1}^{k} \left(\hat{y}^n - (b + w \cdot x^n)\right)^2$$

Step 3: Best Function



Model

$f_1, f_2 \cdots$

$$L(w,b) = \sum_{n=1}^{10} \left(\hat{y}^n - \left(b + w \cdot x_{cp}^n\right)\right)^2$$

Pick the "Best" Function

$$f^* = arg \min_f L(f)$$

$$w^*, b^* = arg \min_{w,b} L(w,b)$$

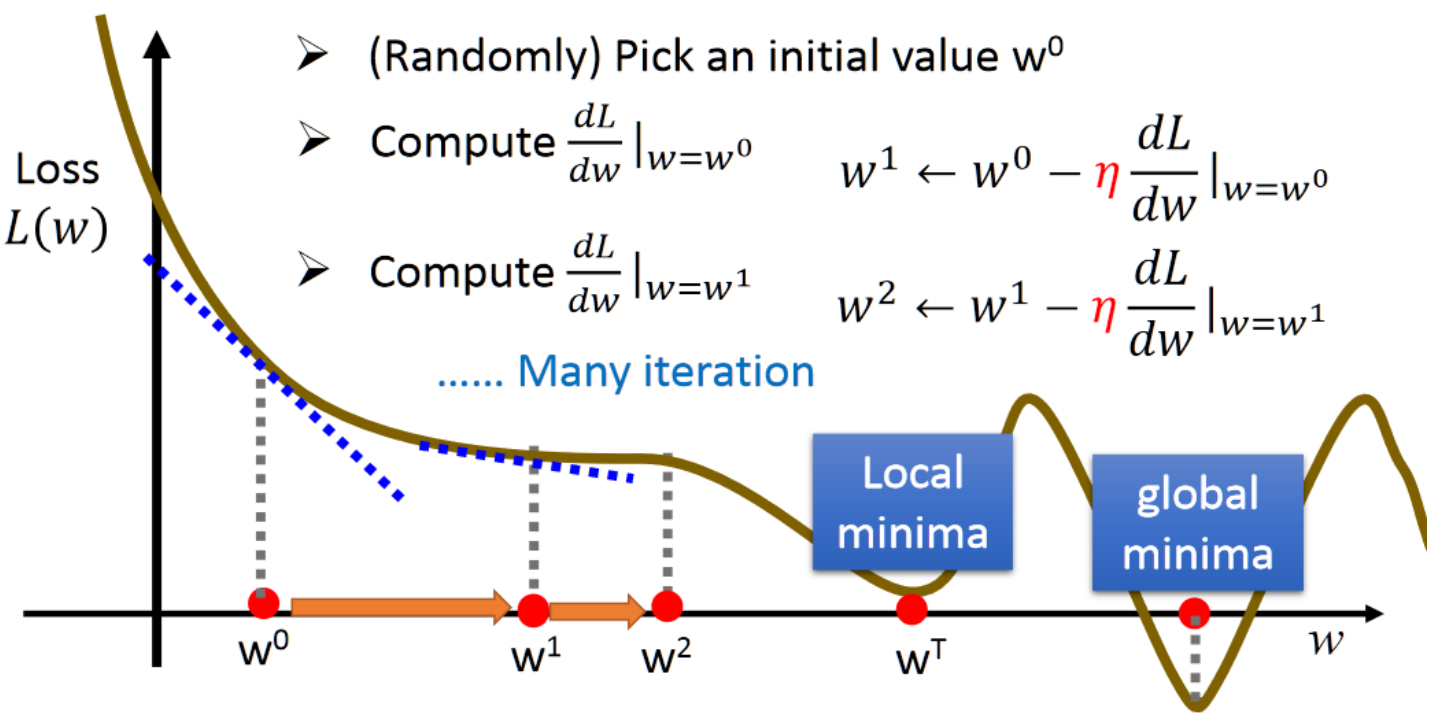
$$= arg \min_{w,b} \sum_{n=1}^{k} \left(\hat{y}^n - (b + w \cdot x^n)\right)^2$$

Gradient Descent

Gradient Descent

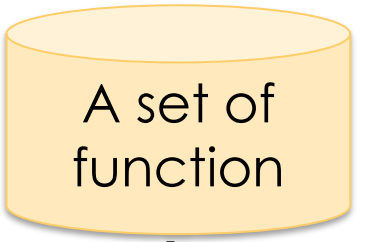
$$w^* = arg \min_w L(w)$$

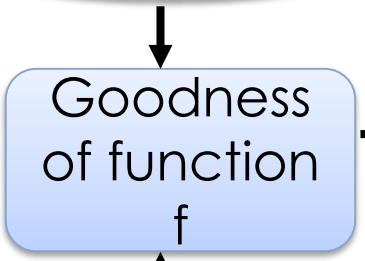
- Consider loss function $L(w)$ with one parameter w:
  - (Randomly) Pick an initial value w^0
  - Compute $\frac{dL}{dw}|_{w=w^0}$ $\quad w^1 \leftarrow w^0 - \eta \frac{dL}{dw}|_{w=w^0}$
  - Compute $\frac{dL}{dw}|_{w=w^1}$ $\quad w^2 \leftarrow w^1 - \eta \frac{dL}{dw}|_{w=w^1}$
  - ..... Many iteration

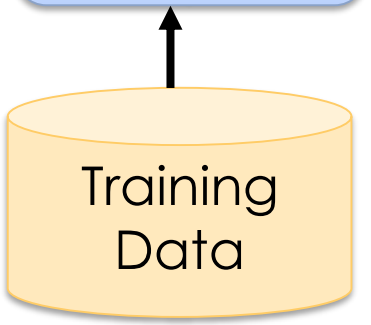


http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017/Lecture/Regression.pdf

Step 3: Best Function



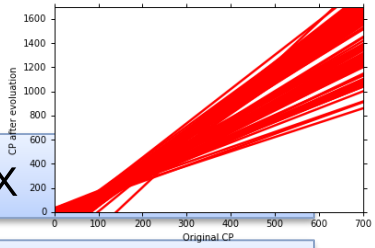




the "Best" Function

What we really care about is the error on new data (testing data)

Something You Should Know



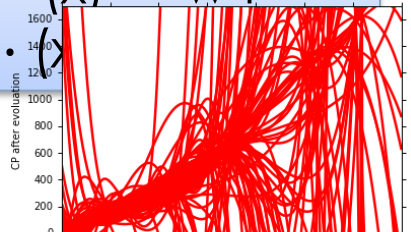
$y = b + w \cdot x$

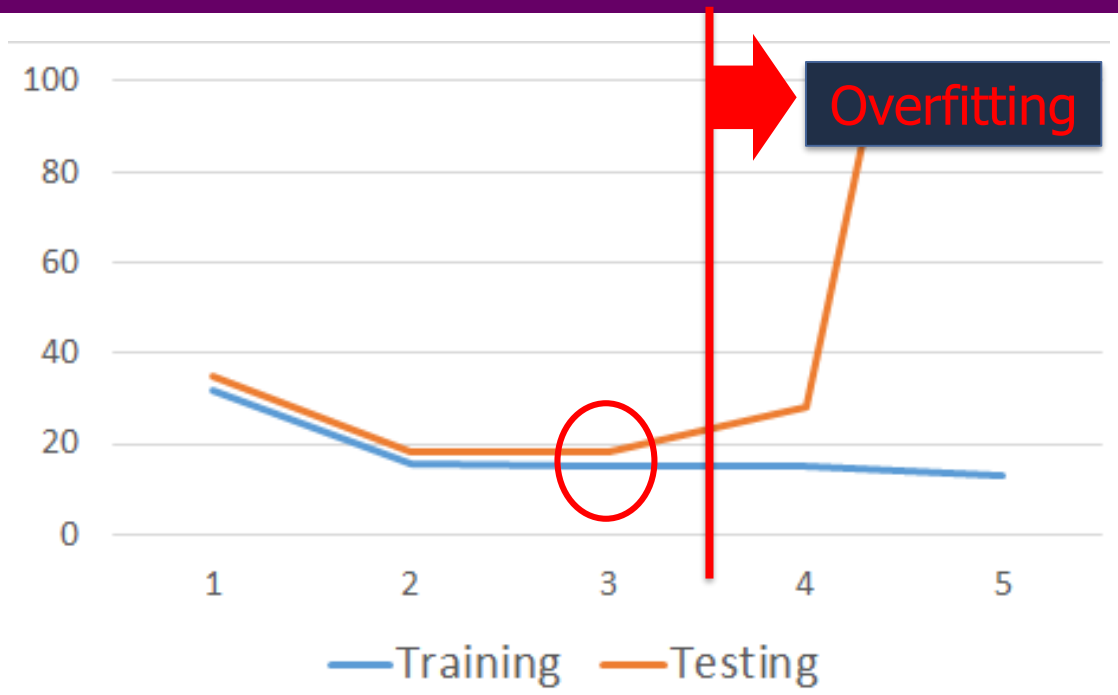
$y = b + w_1 \cdot x + w_2 \cdot (x)^2$

$y = b + w1 \cdot x + w2 \cdot (x)^2$

$y = b + w1 \cdot x + w2 \cdot (x)^2 + w3 \cdot (x)^3 + w4 \cdot (x)^4$

$y = b + w1 \cdot x + w2 \cdot (x)^2 + w3 \cdot (x)^3 + w4 \cdot (x)^4 + w5 \cdot (x$

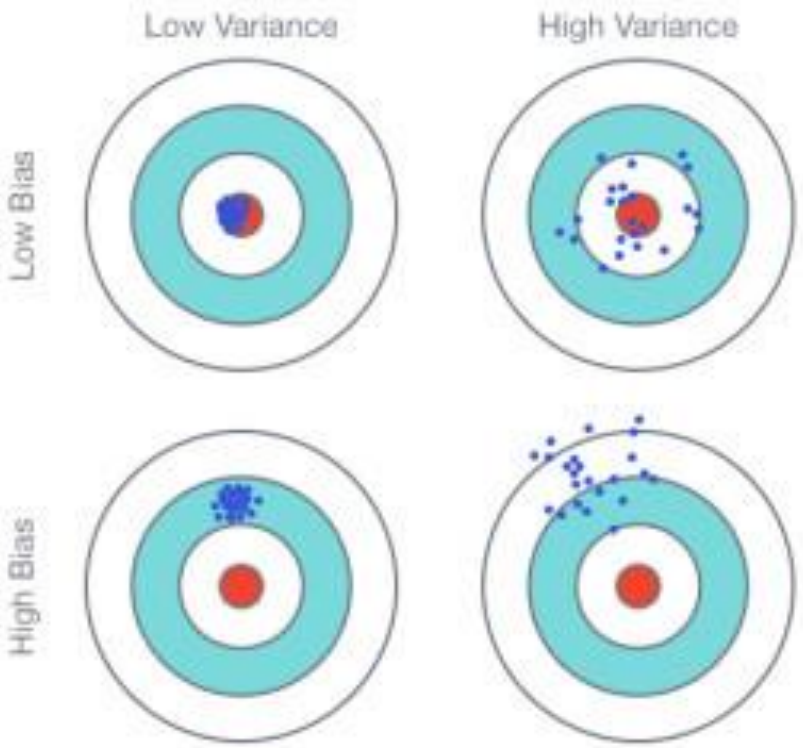




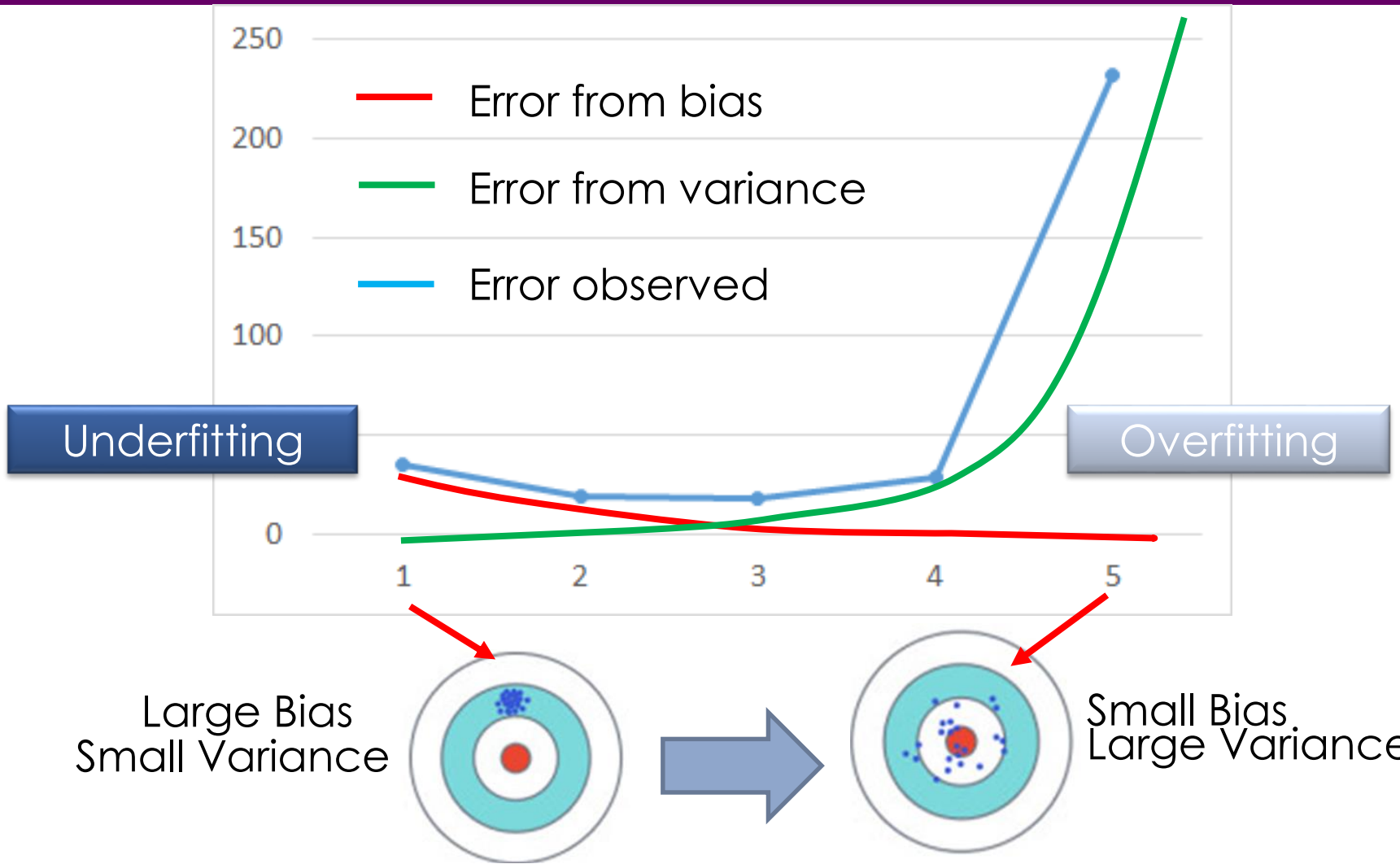
A more complex model yields lower error on **training data**.
However, it **does not always** lead to better performance on **testing data.**

Bias and Variance

Bias v.s. Variance



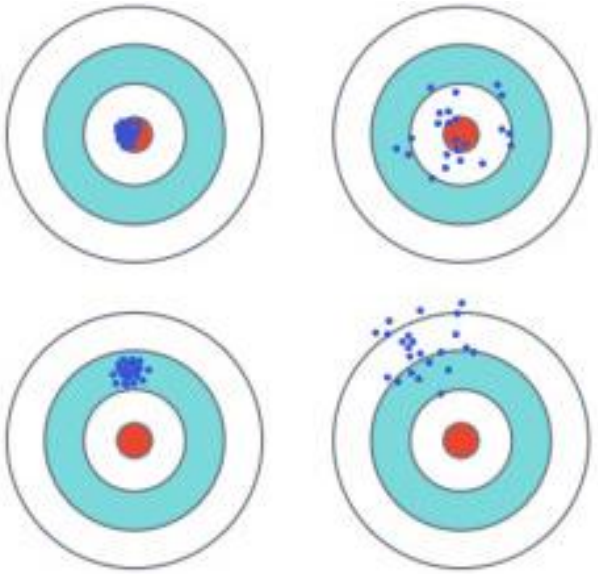
Overfitting, Underfitting



What to do with large bias or variance?

- For bias, redesign your model:
  - Add more features as input
  - A more complex model

- For Variance
  - more data：
  - Very effective, but not always practic
  - Regularization



國立清華大學 National Tsing Hua University

INNOVATION ENGINEERING LAB

References

台大李宏毅老師：

- http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017/Lecture/Regression.pdf
- http://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2017/Lecture/Bias%20and%20Variance.pdf

Classroom Test

- Please use The Housing dataset from UCI to predict the house prices.

  You are asked to:

  1. generate a pairwise plot to observe how the data is distributed, and briefly illustrate what you have observed
  2. show your linear regression parameters(slope and bias)
  3. visualize how the linear regression line fits the training data

- Turn in your work with the format of .ipynb , and please write some brief comments in your ipynb to illustrate your results.

Homework

- Please use the 'insurance.csv' file on iLMS and what we taught in TA class to predict the insurance charge with age.

  You are asked to:

  1. generate a pairwise plot to observe how the data is distributed, and briefly illustrate what you have observed, you may add other plot to observe data
  2. show your linear regression parameters(slope and bias)
  3. visualize how the linear regression line fits the data

➢ Hint: you can use 'LabelEncoder()' to transform categorical data to label

- Turn in your work with the format of .ipynb , and please write some brief comments in your ipynb to illustrate your results.