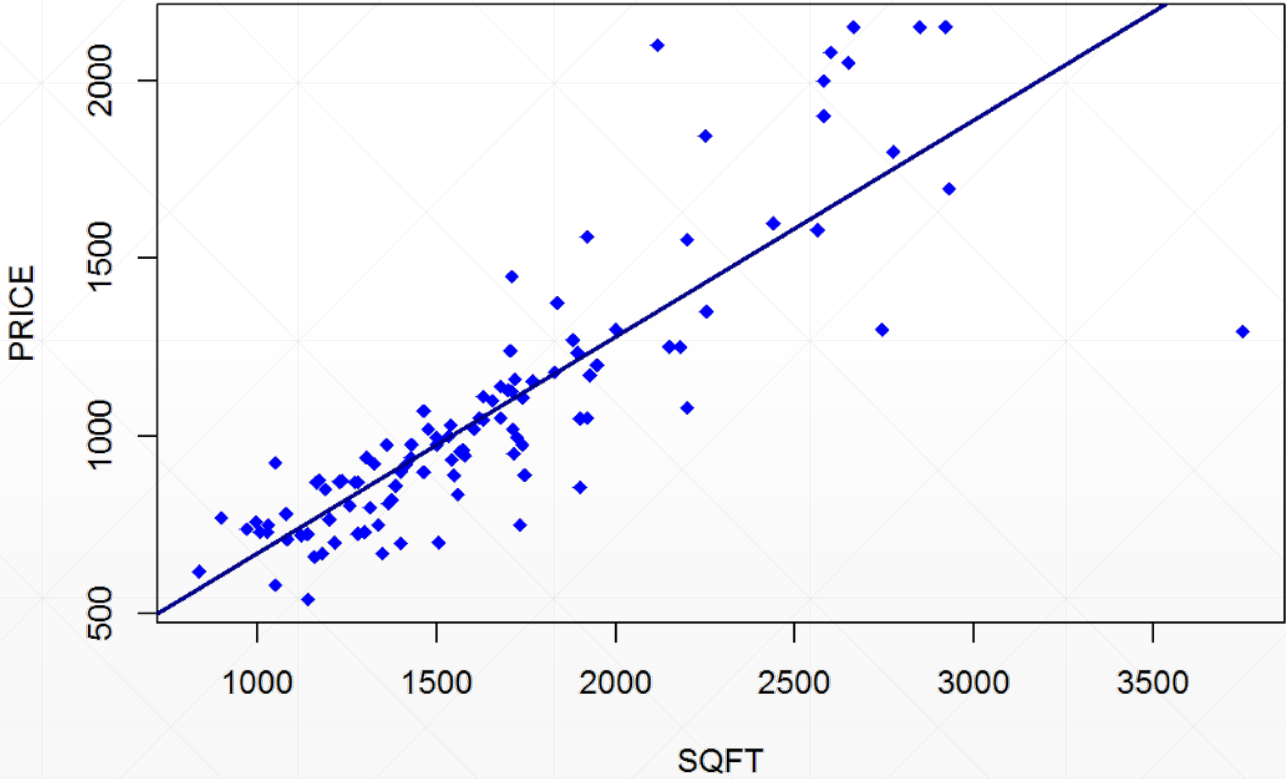


过拟合与欠拟合

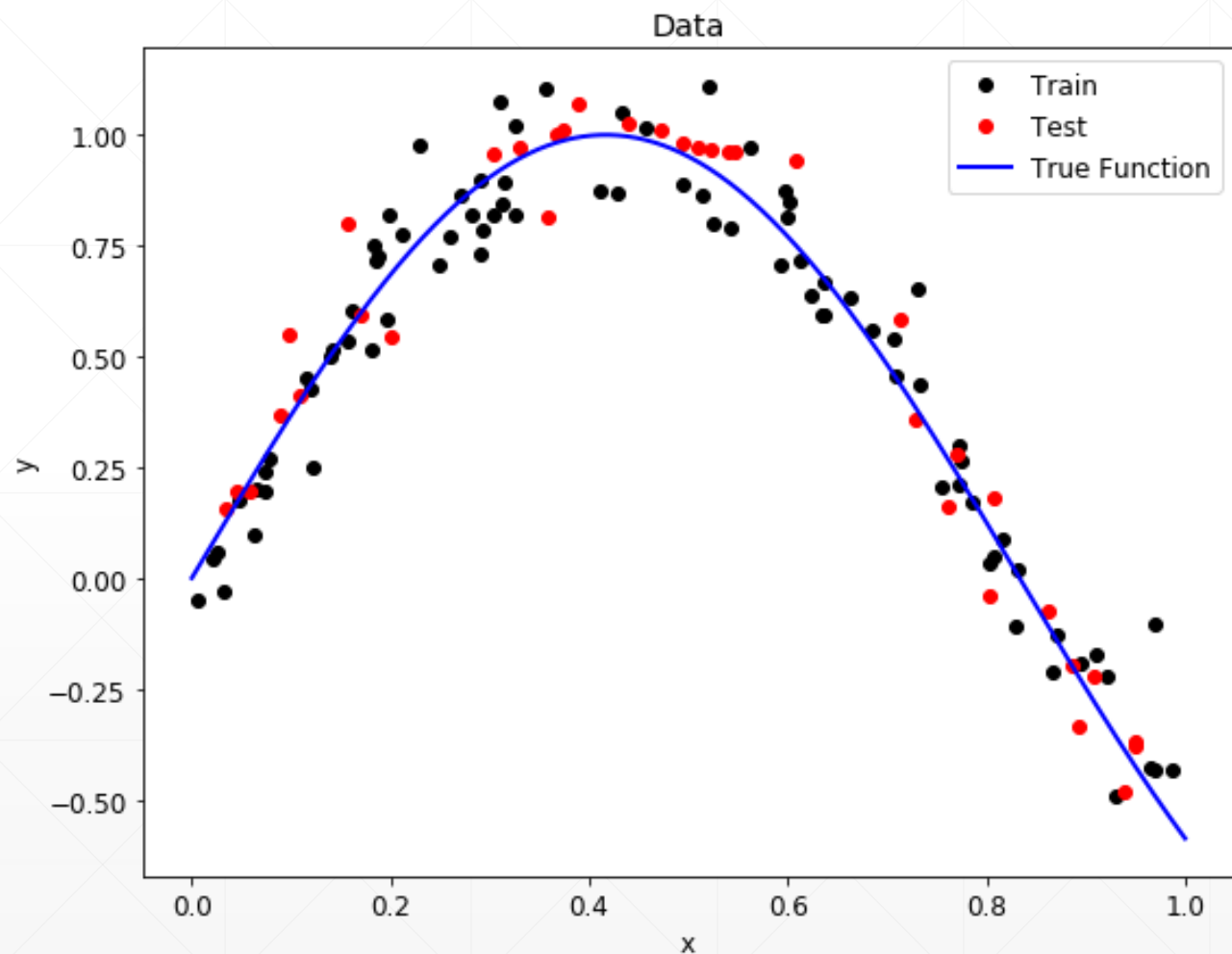
主讲：龙良曲

Scenario 1: House Price

Price vs Square Footage (with regression line)

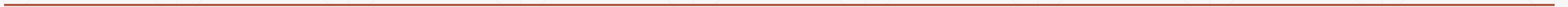


Scenario2: GPA



Ground-truth distribution?

- That's perfect if known
- However



Another factor: noise

- $y = w * x + b + \epsilon$
- $\epsilon \sim N(0.01, 1)$

- $1.567 = w * 1 + b + \text{eps}$
- $3.043 = w * 2 + b + \text{eps}$
- $4.519 = w * 3 + b + \text{eps}$
- ...

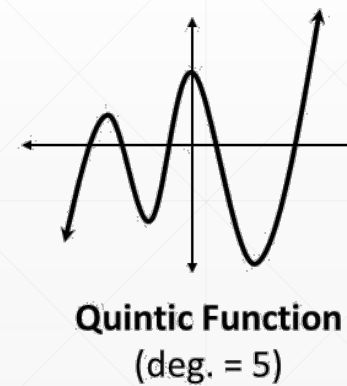
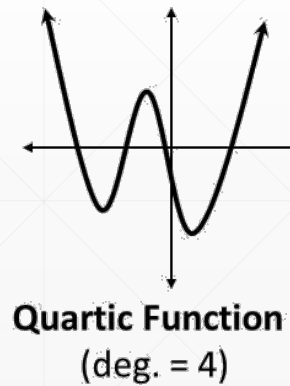
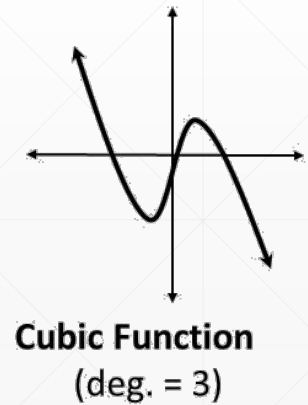
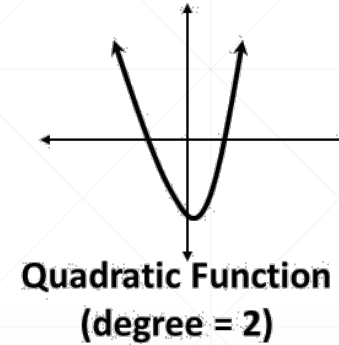
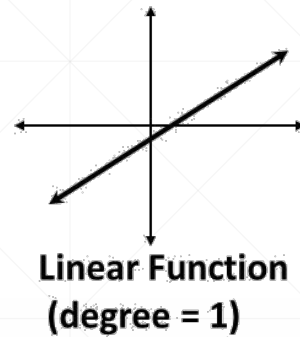
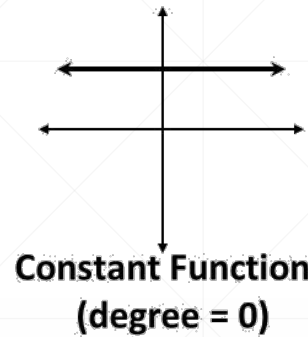
$$\text{loss} = (WX + b - y)^2$$



Let's assume

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \varepsilon.$$

Graphs of Polynomial Functions:



Mismatch: ground-truth VS estimated

- model capacity

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots + \beta_n x^n + \varepsilon.$$

Model Capacity

Revolution of Depth

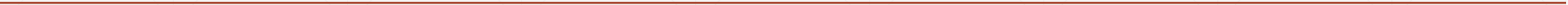
AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)

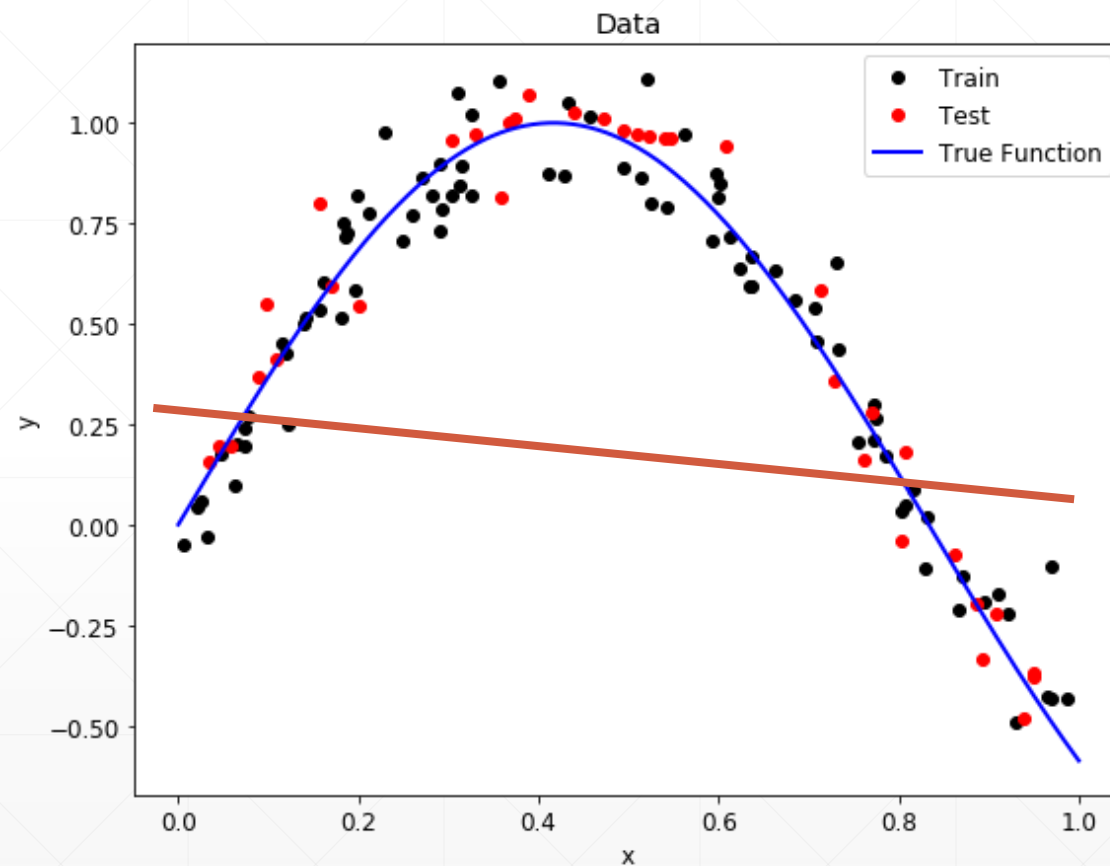


ResNet, 152 layers
(ILSVRC 2015)



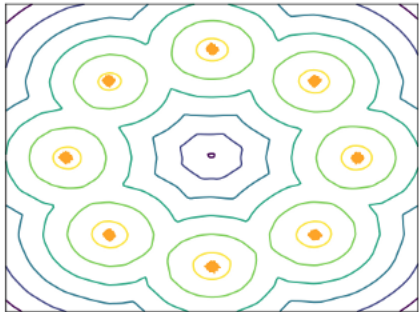
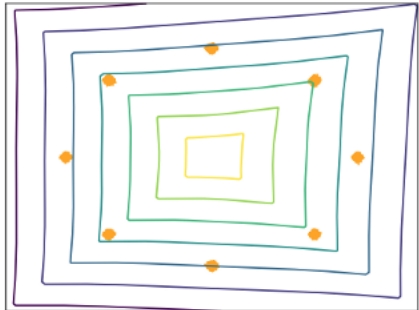
Case1: Estimated < Ground-truth

under-fitting

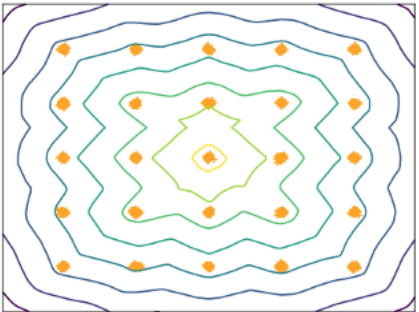
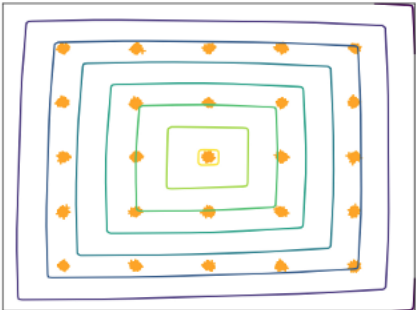


For example: WGAN

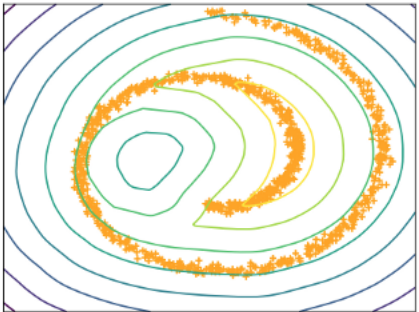
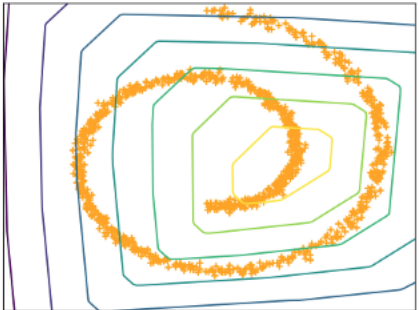
8 Gaussians



25 Gaussians

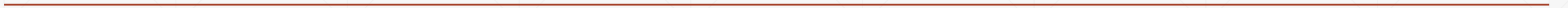


Swiss Roll



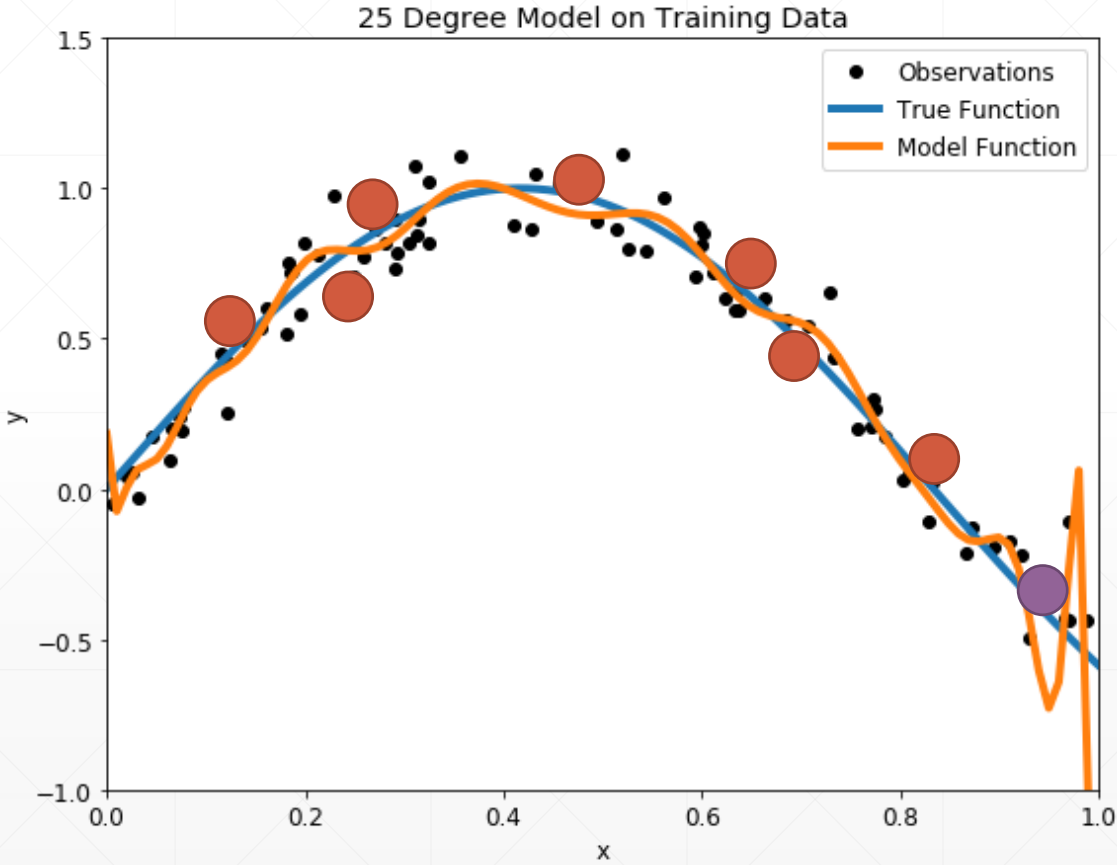
Underfitting

- train acc. is bad
- test acc. is bad as well



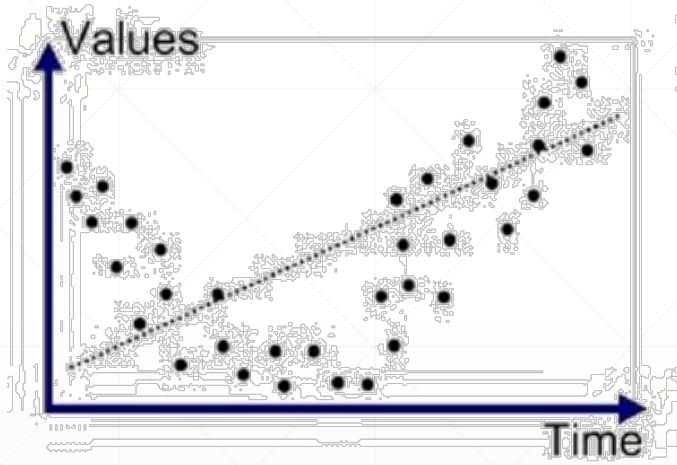
Case2: Ground-truth < Estimated

Over-fitting

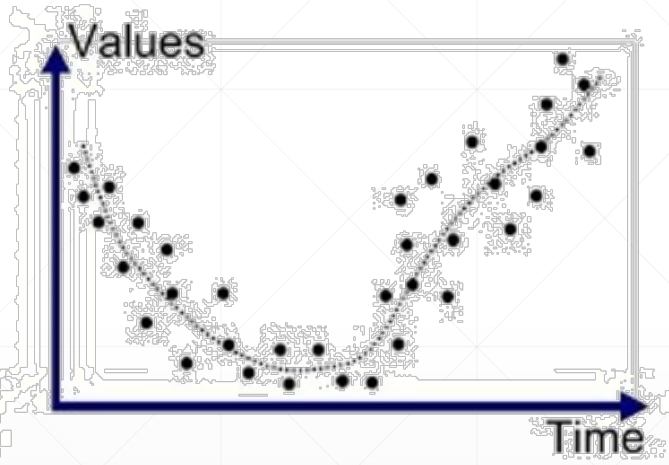


Overfitting

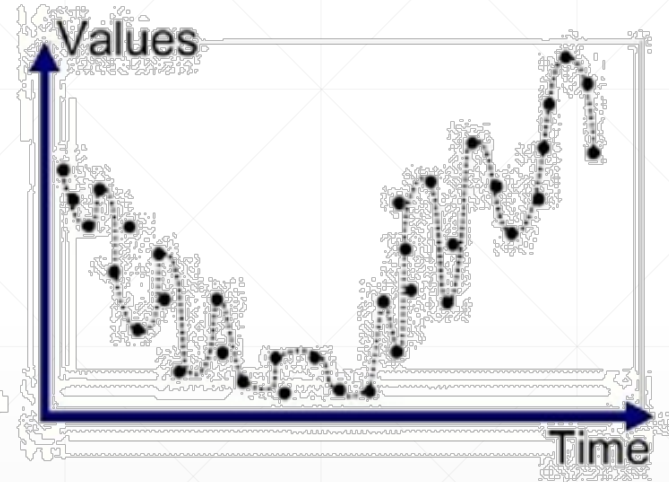
- train loss and acc. is much better
 - test acc. is worse
 - → Generalization Performance
-



Underfitted



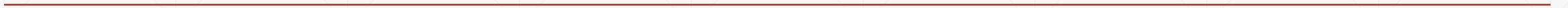
Good Fit/Robust



Overfitted

Overfitting !

- how to detect
- how to prevent



下一课时

交叉验证

Thank You.
