Computational Intelligence

Samaneh Hosseini

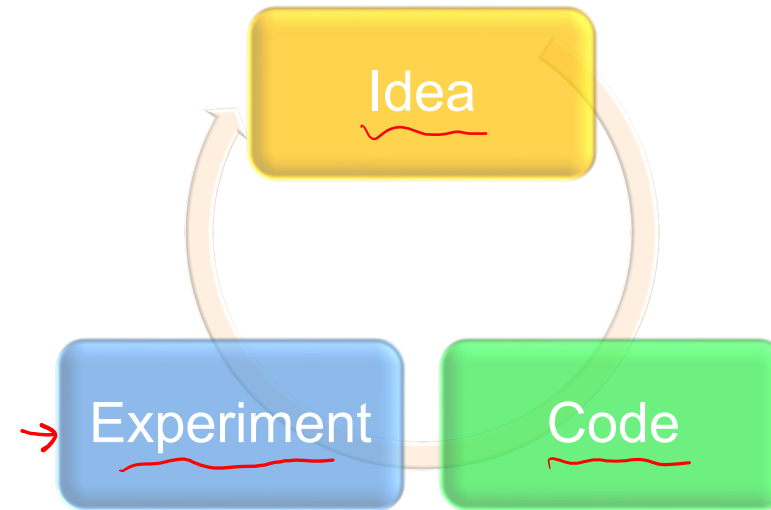Isfahan University of Technology

Outline

- Neural Networks in Practice

  - Train/dev/test sets

  - Overfitting

Neural Networks in Practice: Train/dev/test sets
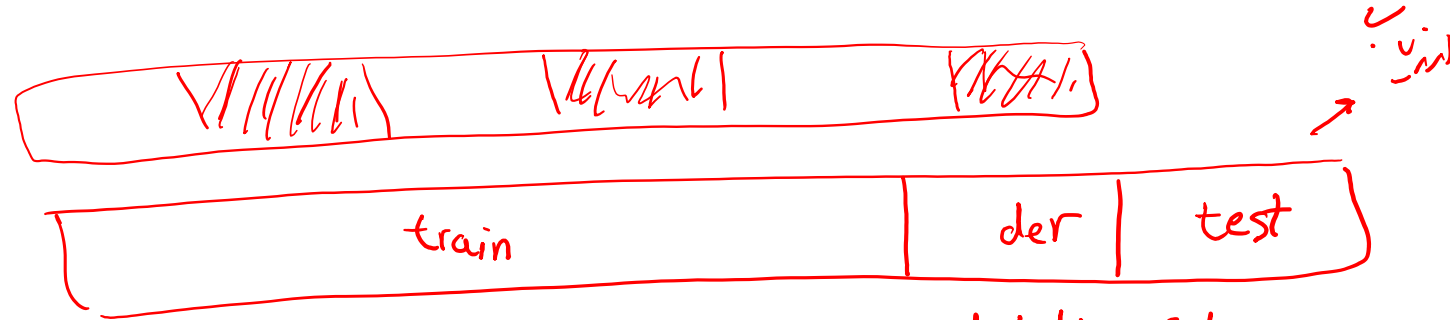
Applied ML is a highly iterative process

hyper parameter

- # layers

- # hidden units

- learning rates

- activation functions

Idea

Experiment   Code

NLP _vision _ Speach _ Structured data

Search ads   Security ...

Train/dev/test sets

validation set
development set
dev

آنچه آموزش شبکه

train dev test

train | der | test

previous data : 70% / 30%

100 — 1000 — 10000

60 / 20 / 20 %

der   test
10,000   10,000

Big data :   1,000,000   der   test

98% / 1% / 1%

99.5% / .25 / .25 %

Mismatched train/test distribution

Training set:
Cat pictures from
webpages

Dev/test sets:
Cat pictures from
users using your app

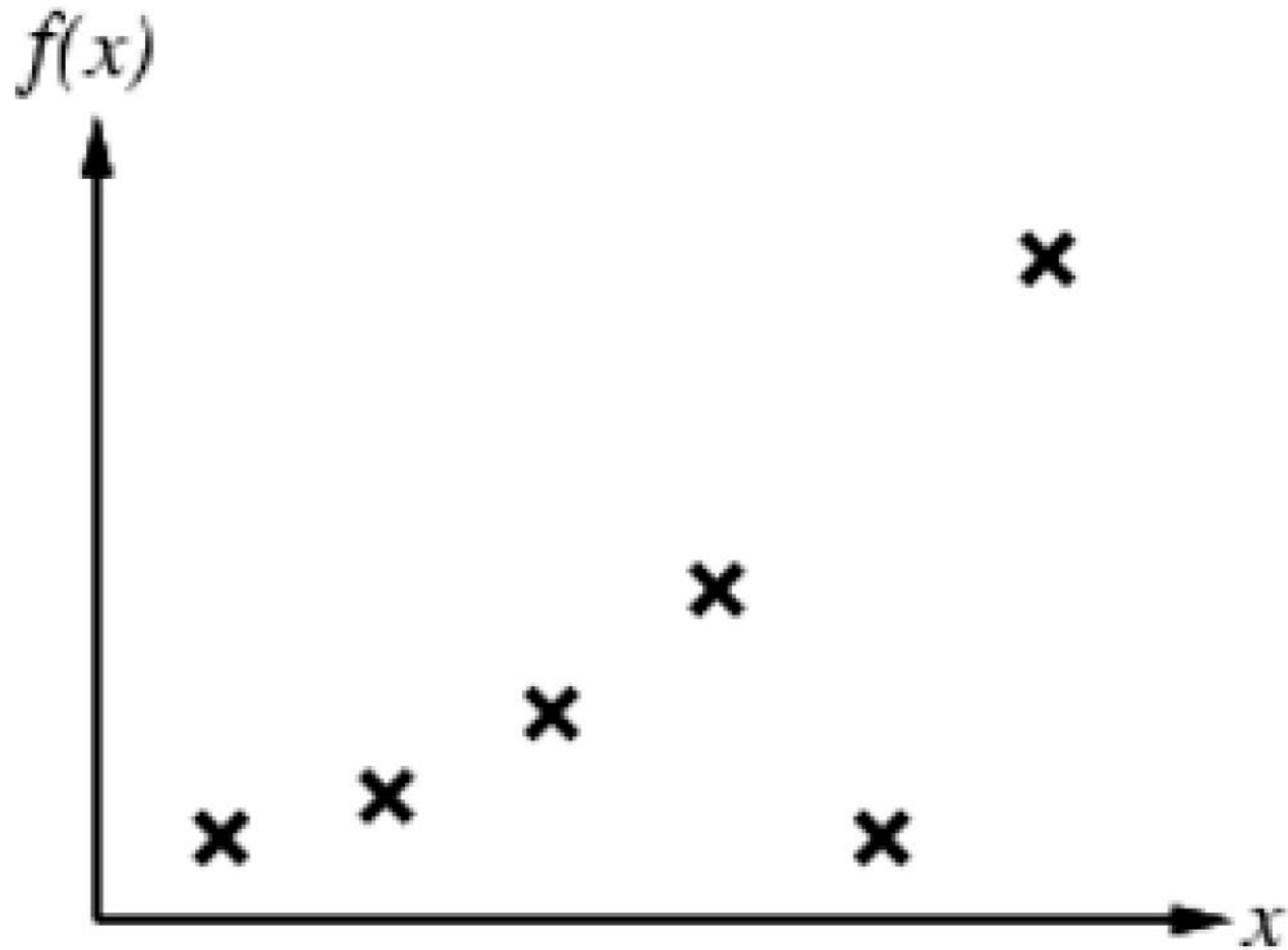Dev و test از یک distribution باشد ✓

train / test ✗   train / dev ✓

Not having a test set might be okay. (Only dev set.)
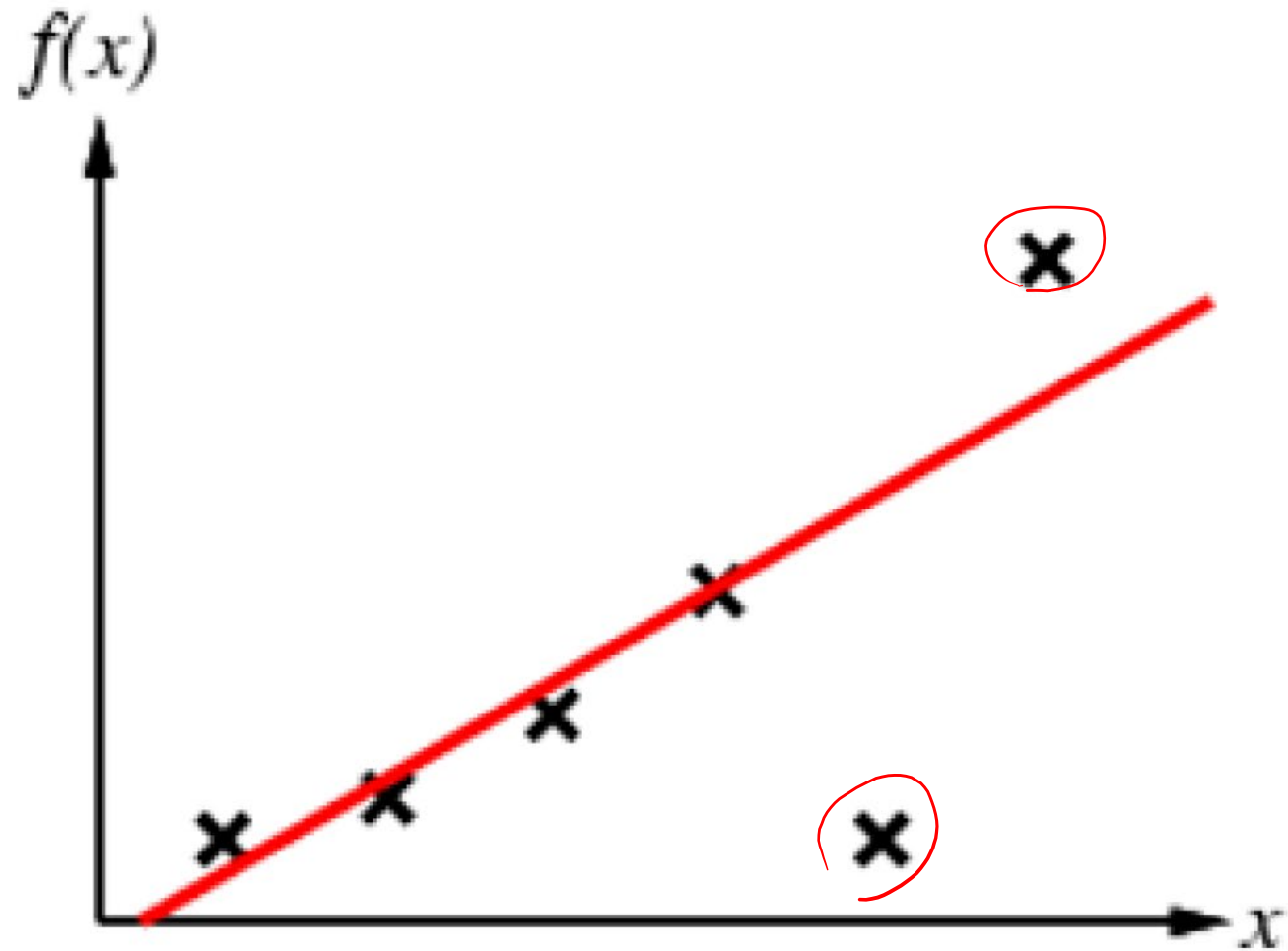
Neural Networks in Practice: Overfitting

bias / variance
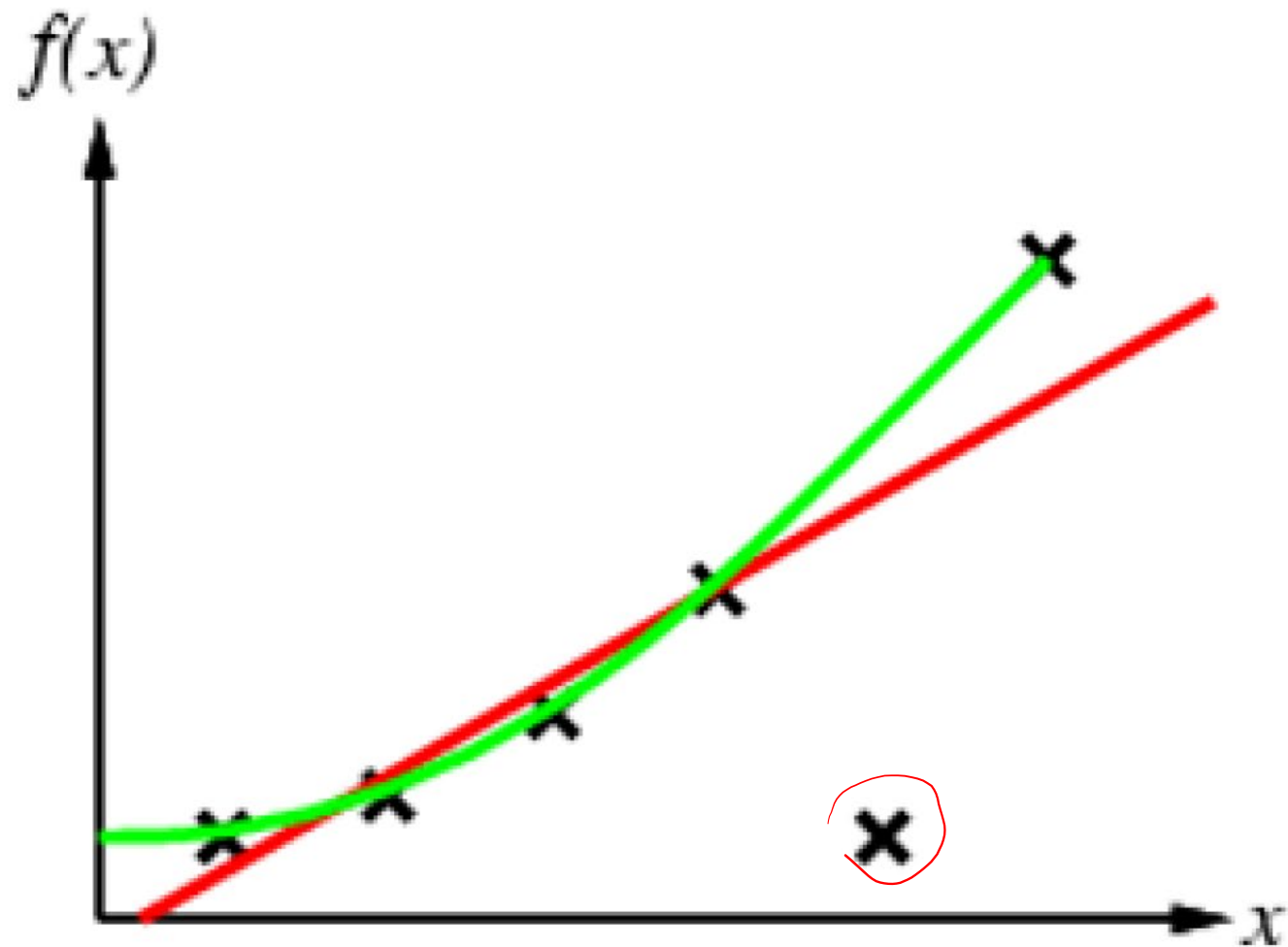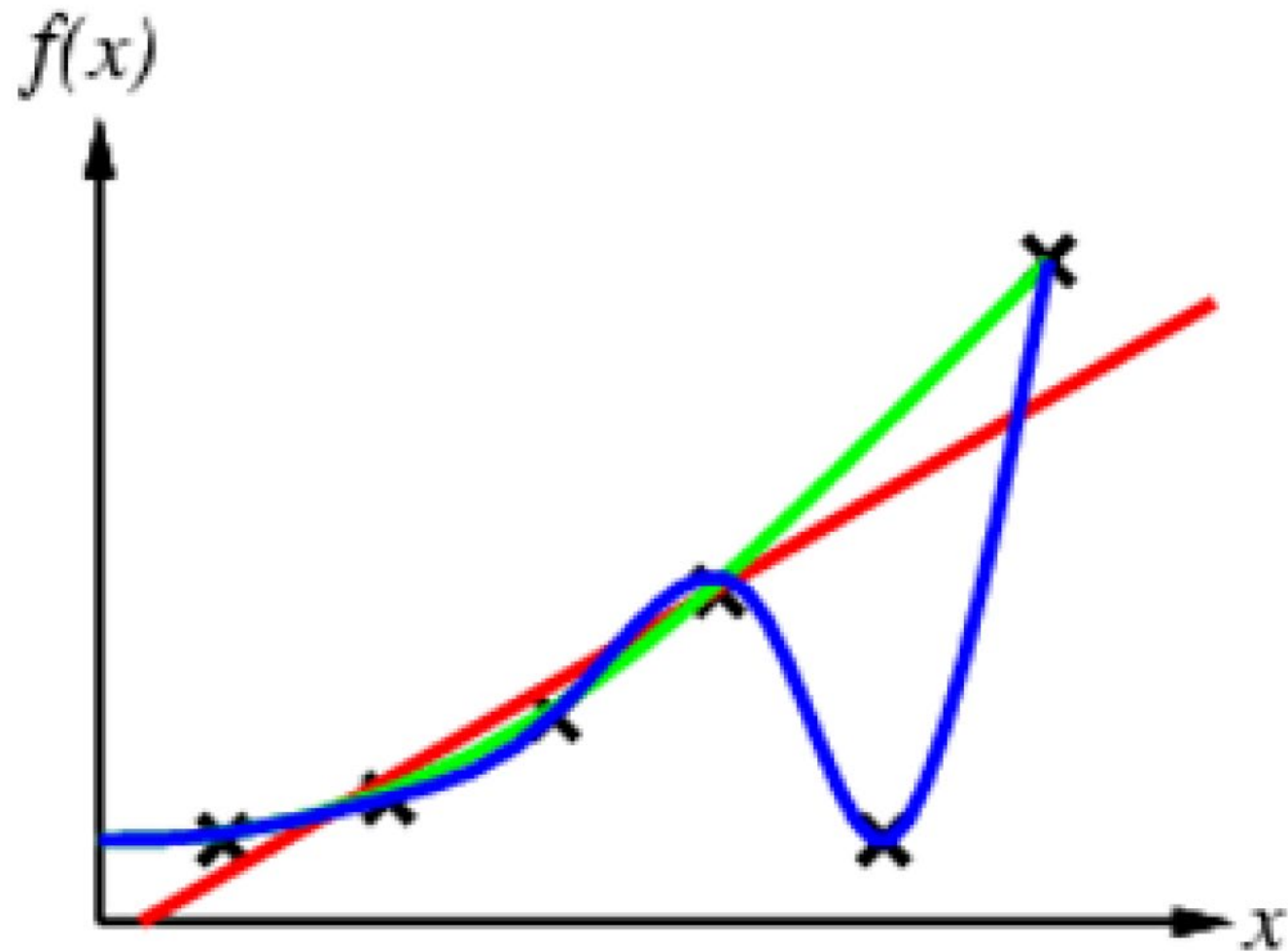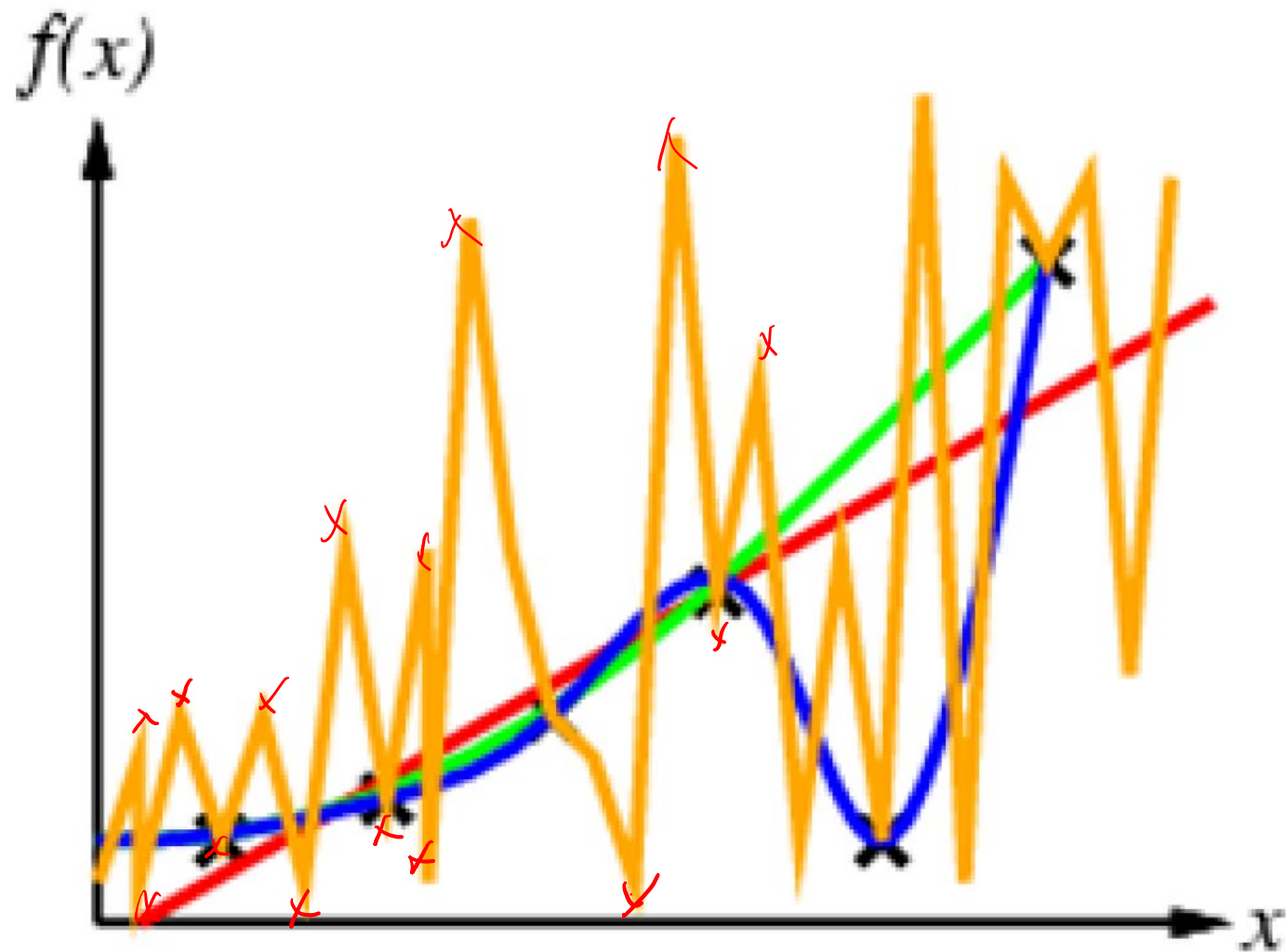
Example: A prediction task

Example: A prediction task
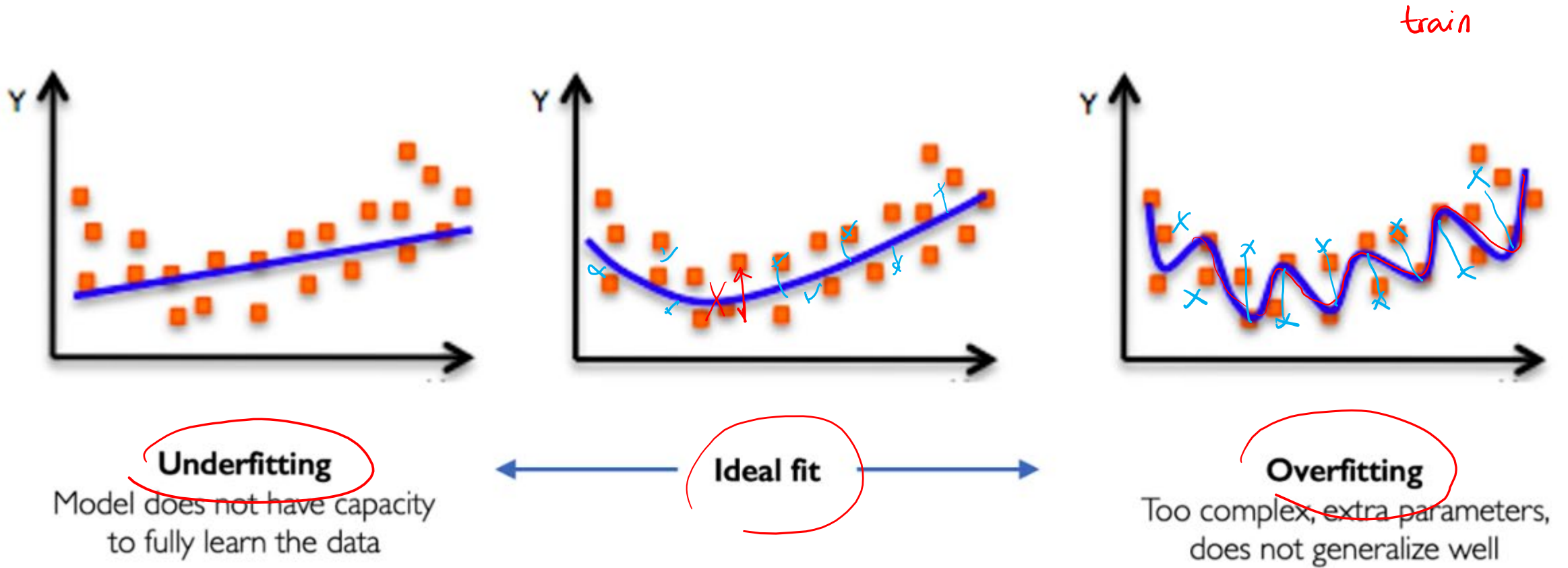
Example: A prediction task
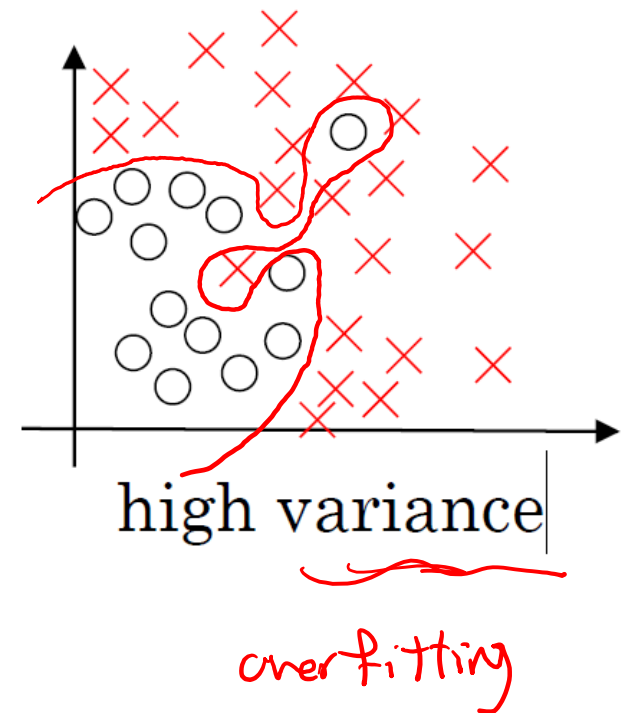
Example: A prediction task
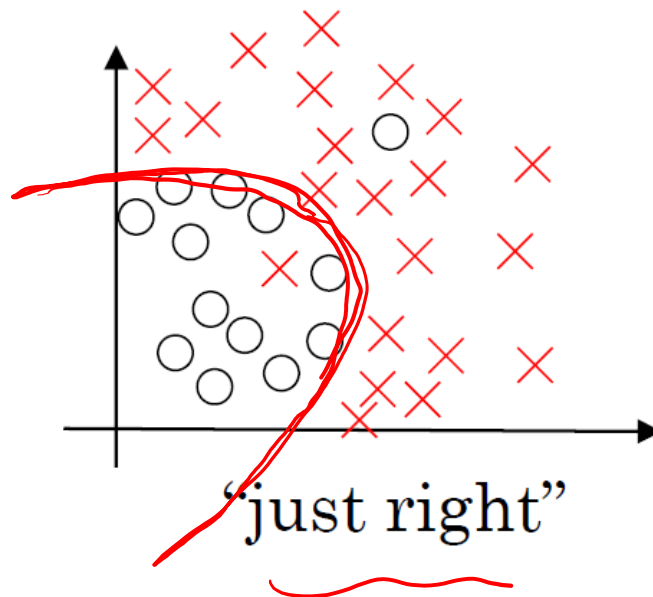
Example: A prediction task

The Problem of Overfitting

Underfitting
Model does not have capacity to fully learn the data

Ideal fit

Overfitting
Too complex, extra parameters, does not generalize well

Bias and Variance

high bias — under fitting

"just right"

high variance — overfitting

Bias and Variance

Cat classification

$y = 1$ $y = 0$

Train set error: → 1%   15%   15%   0.5%

Dev set error: → 11%   16%   30%   1%

high variance   high bias   high bias   low bias
                              high variance   low variance
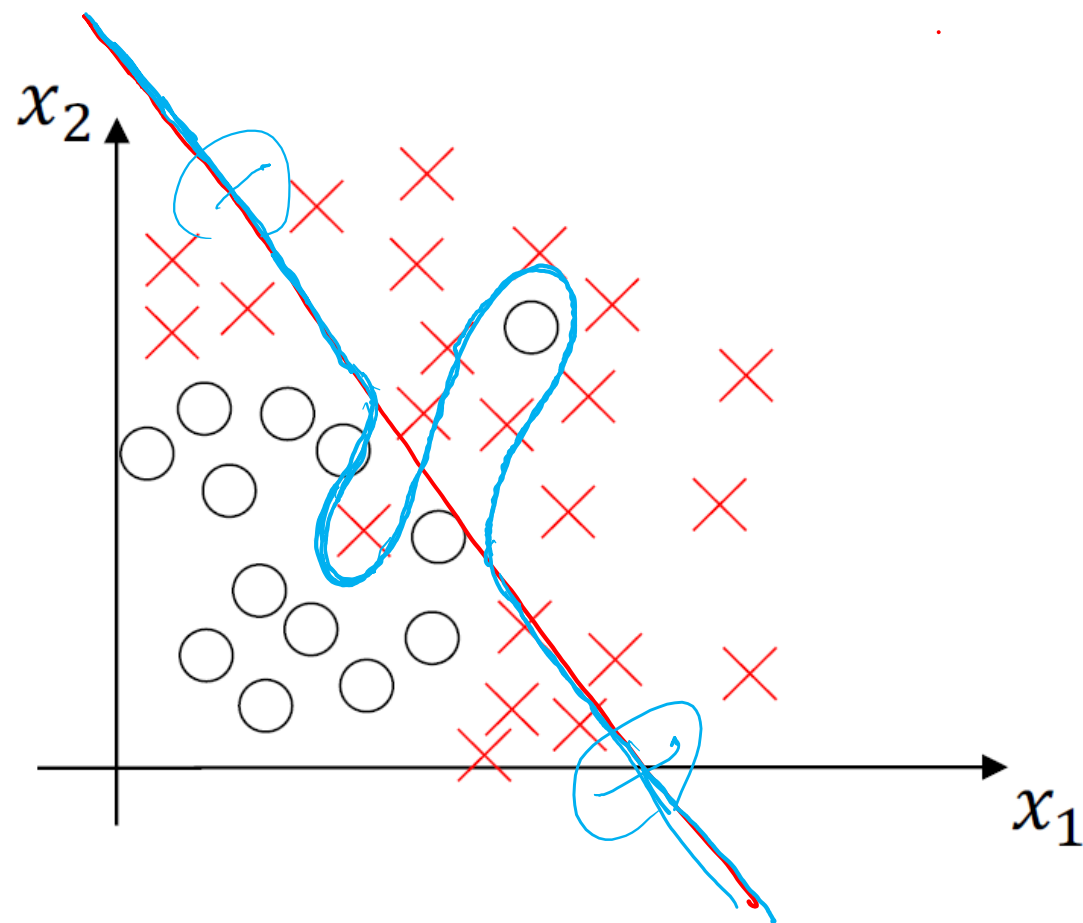
→ human error : 0%   15%

→ Optimal (Bayes) error ≃ 0%
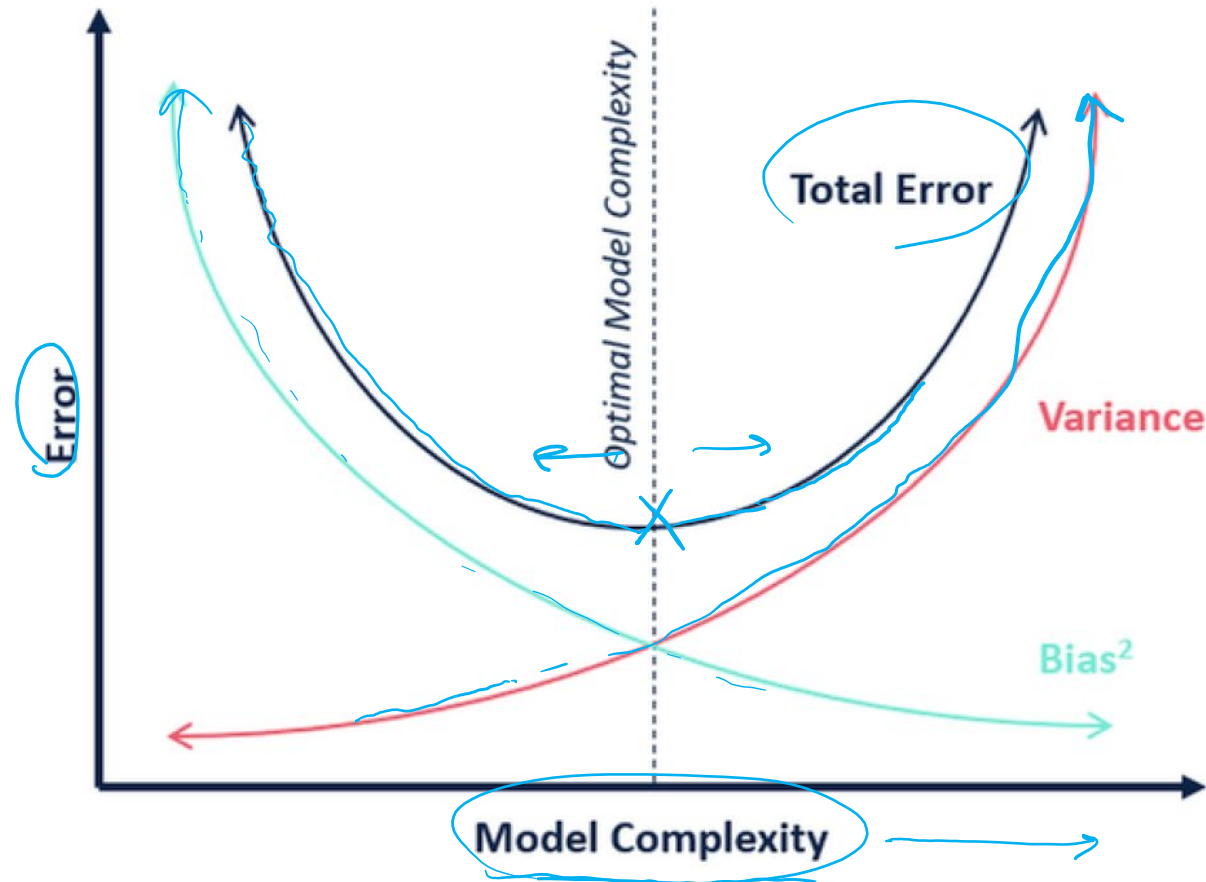
High bias and high variance

high bias
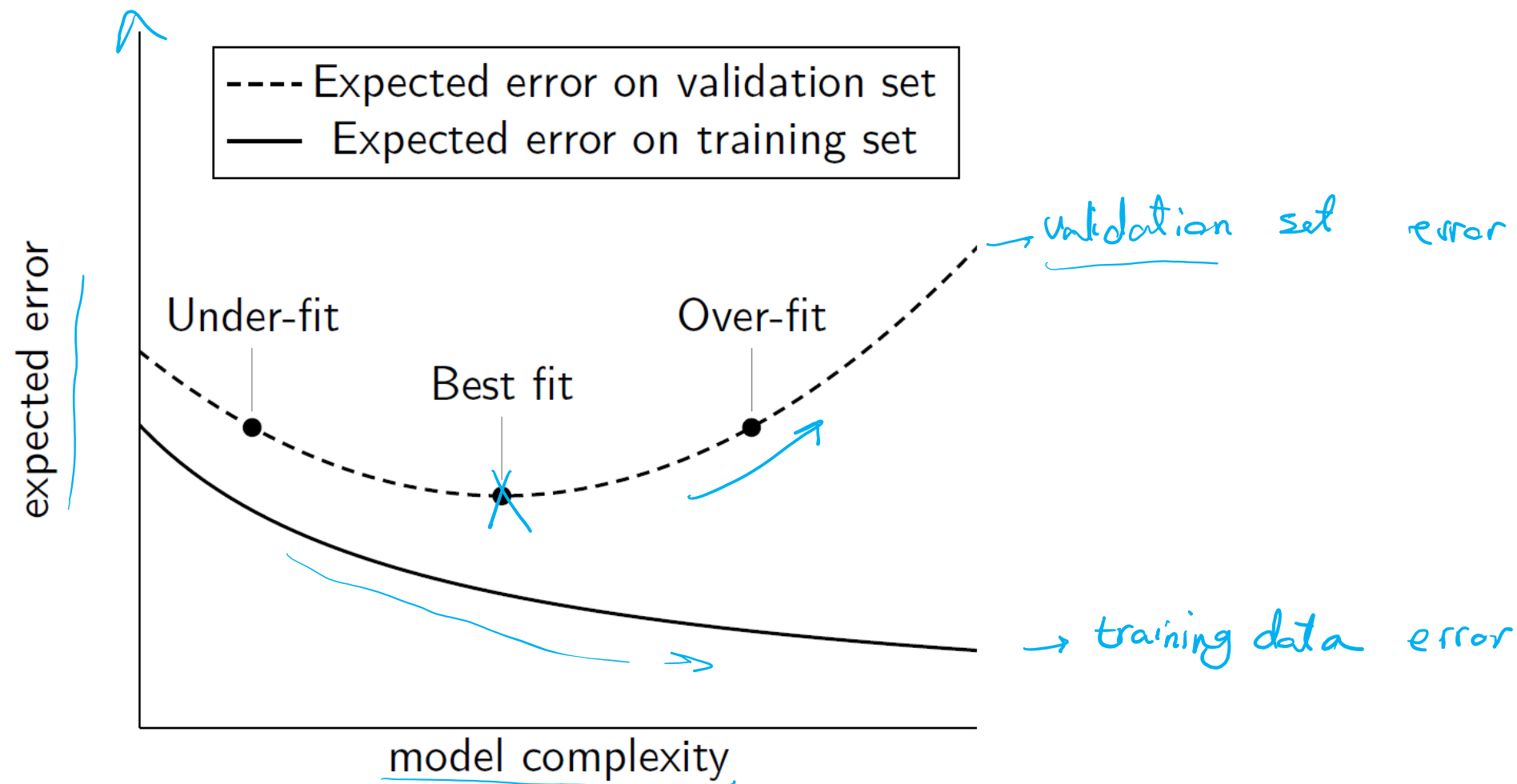high variance

Overfitting: The Problem of Generalization

▶ Goal of ML is to find a hypothesis
that can predict unseen examples correctly.

▶ A trade-off between
  ▶ complex hypotheses that fit the training data well
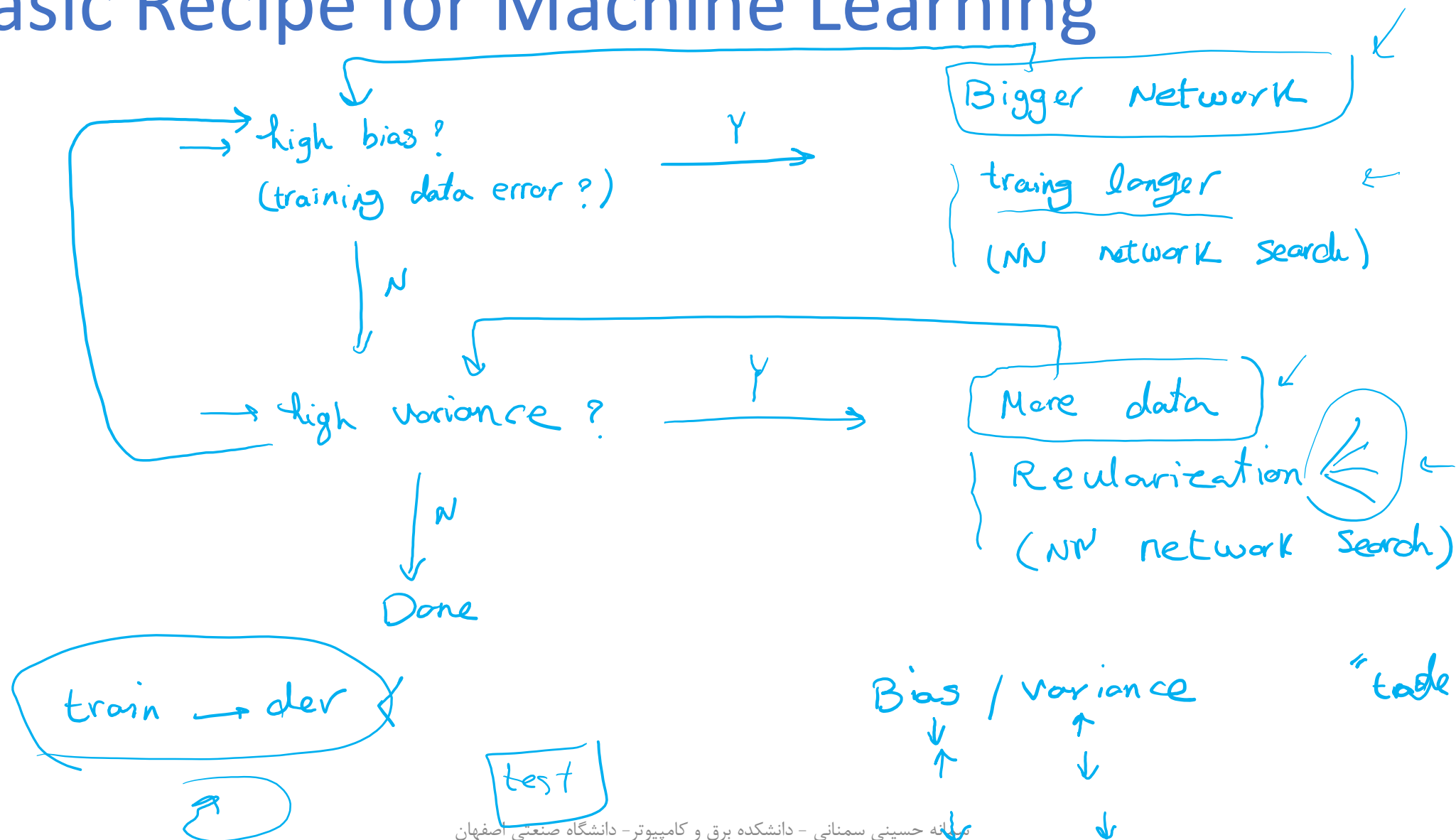 ▶ simpler hypotheses that may generalize better

Bias-Variance Trade-off

How well does the hypothesis fit the data
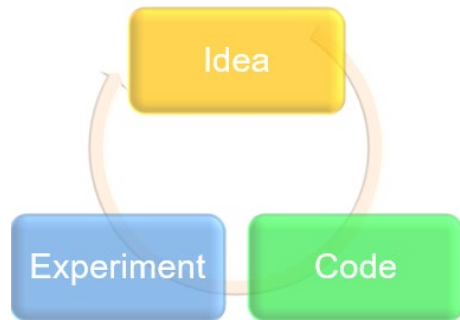as the hypothesis becomes more complex?

Over-fitting

Expected error on validation set — — — —

Expected error on training set ————

validation set error

Under-fit

Best fit

Over-fit

expected error

training data error

model complexity

Basic Recipe for Machine Learning

high bias ?
(training data error ?)

Y → Bigger Network

traing longer

(NN network search)

N

high variance ?

Y → More data

Reularization

(NN network Search)

N

Done

train → dev

test

Bias / variance

"tade all"
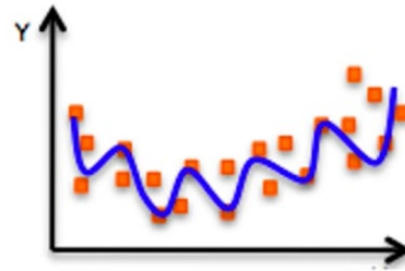
Core Foundation Review

Iterative process

Overfitting

Bias/Variance tradeoff

Overfitting
Too complex, extra parameters,
does not generalize well