Welcome to

DS3010:
DS-III: Computational Data Intelligence
Bias and Variance

Prof. Yanhua Li

Time: 11:00am – 12:50pm M & R
Location: HL 114
D-term 2022

Quiz #1

- *Note:* Quiz 1 on Canvas
- Week 4 (4/4 M)
- 15 mins at the beginning of the class.
- Topics: Bias and Variance

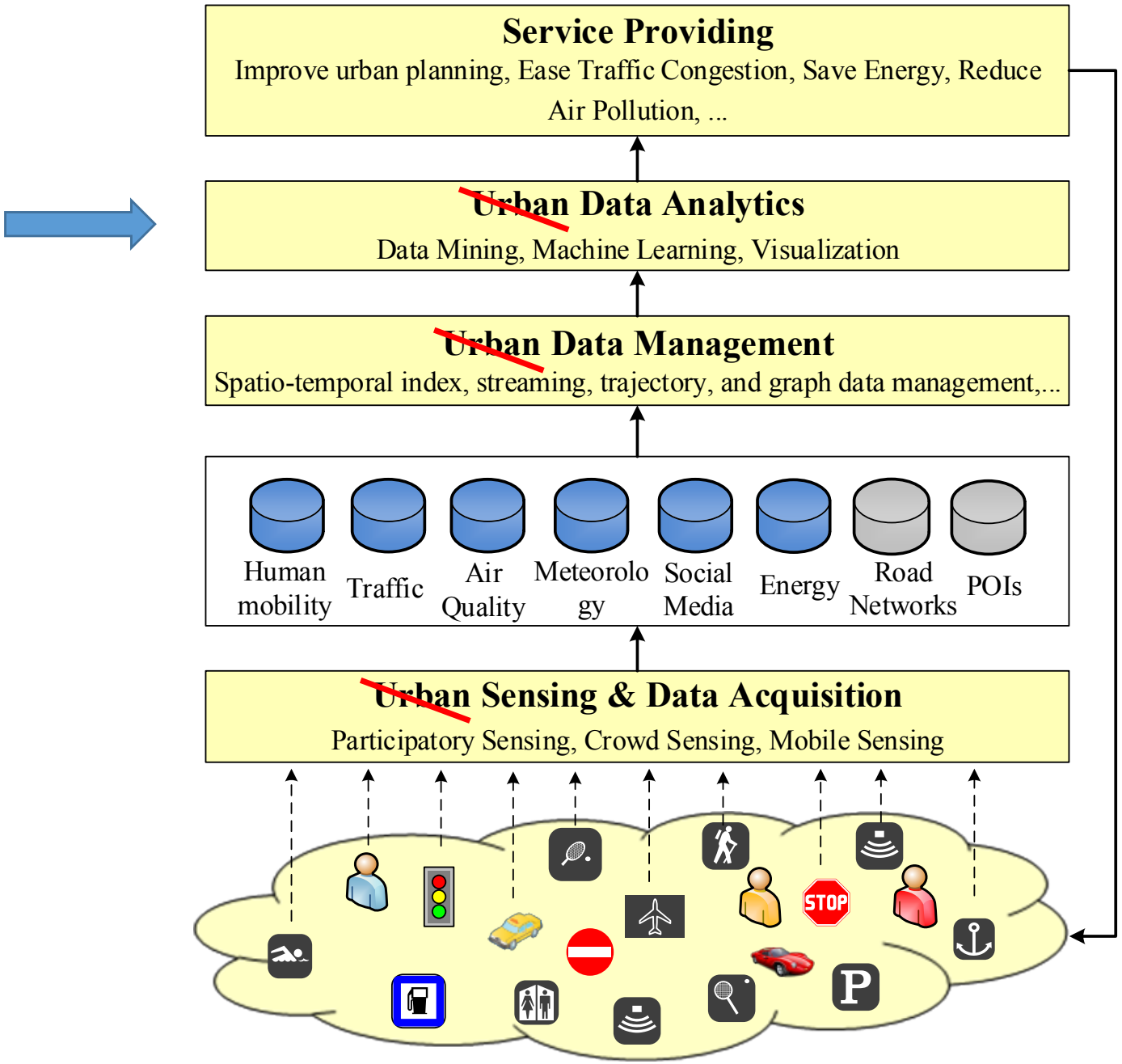
Project #1

- *Due on* Week 3 (3/31 R)
- Submitted it on Canvas
- https://github.com/ds3010s22/ds3010_projects/blob/main/Project1_Twitter.ipynb

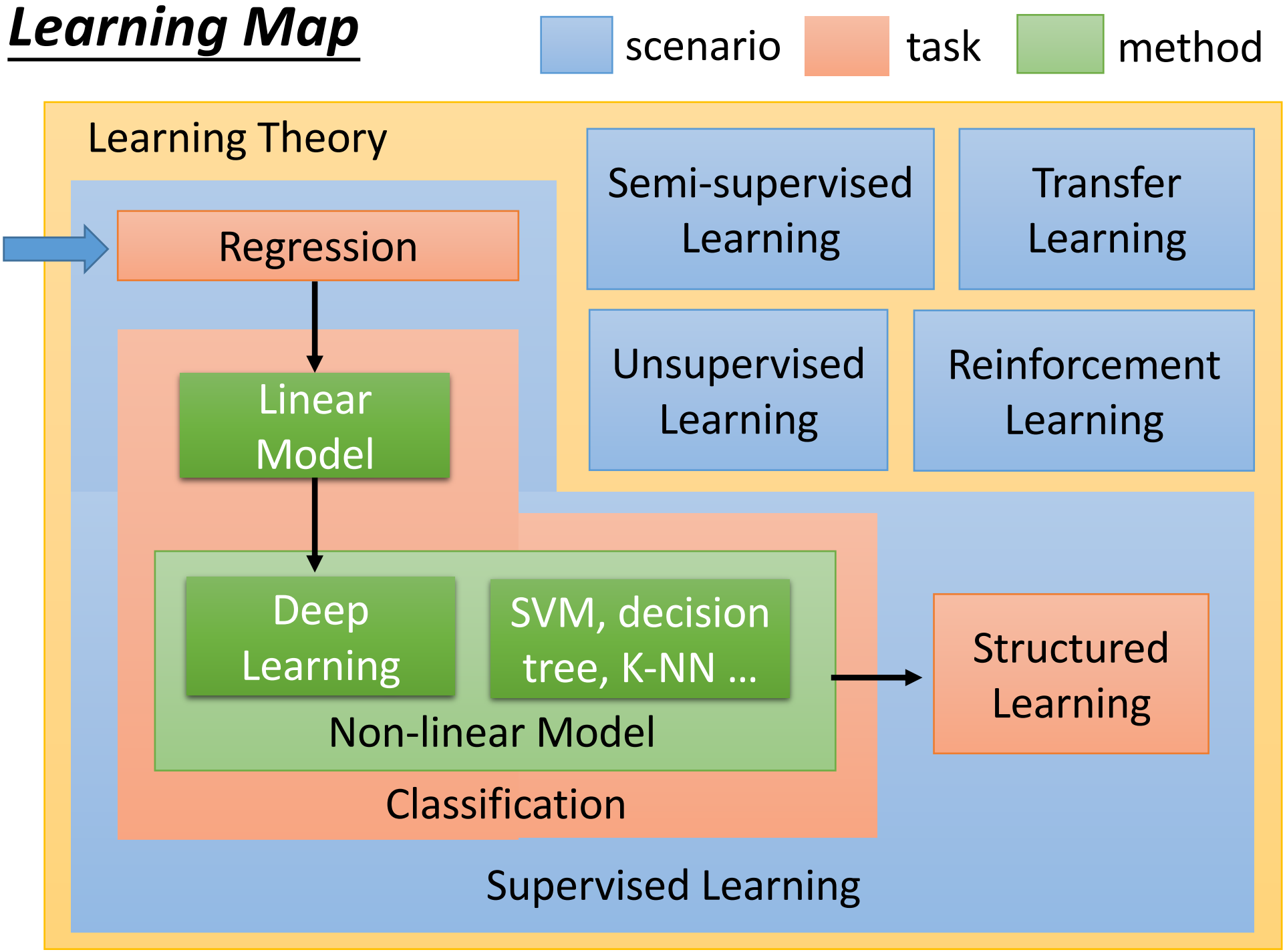
Project #2 (Data Analytics and Machine Learning)

- *Starts on* Week 3 (3/31 R)
- *Due on* Week 6 (4/18 M)
- **Analysis of Mobile Phone Price/Cost**
- Submitted it on Canvas
- https://github.com/ds3010s22/ds3010_projects/blob/main/Project2.ipynb

Data pipeline

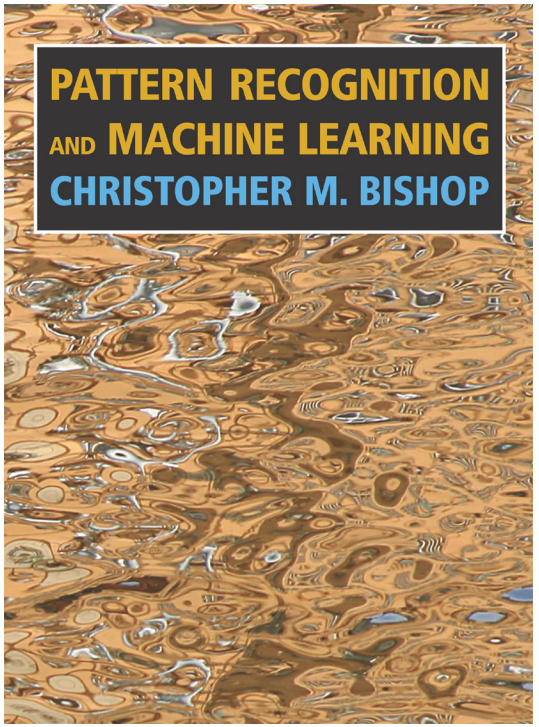


Urban Computing: concepts, methodologies, and applications.
Zheng, Y., et al. *ACM transactions on Intelligent Systems and Technology*.


Learning Map
scenario
task
method
Learning Theory
Semi-supervised Learning
Transfer Learning
Regression
Unsupervised Learning
Reinforcement Learning
Linear Model
Deep Learning
SVM, decision tree, K-NN …
Non-linear Model
Structured Learning
Classification
Supervised Learning

References

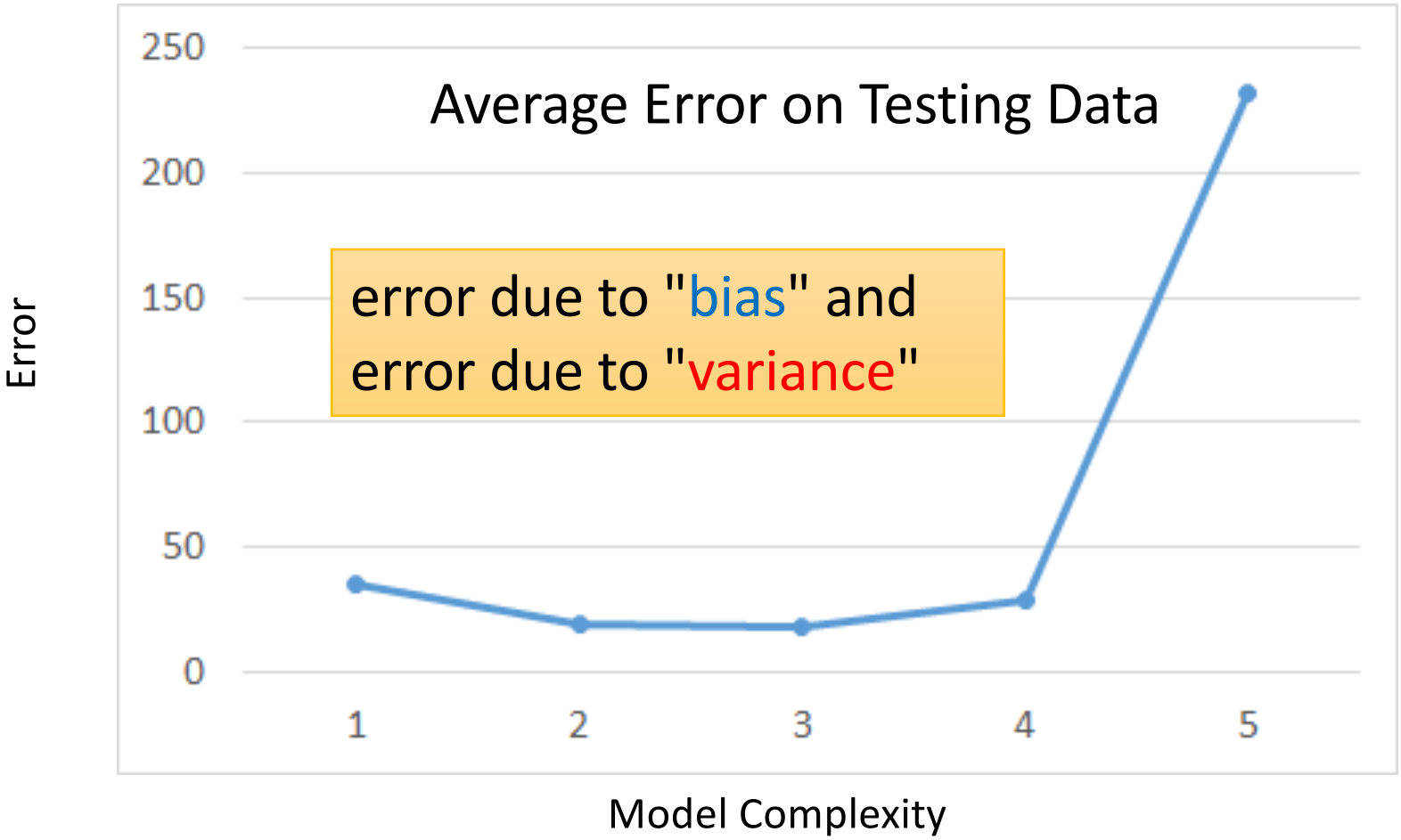
Regression
Bias and variance



Bishop: Chapter 3.2

Where does the error come from?

Review



A more complex model does not always lead to better performance on ***testing data***.

Estimator

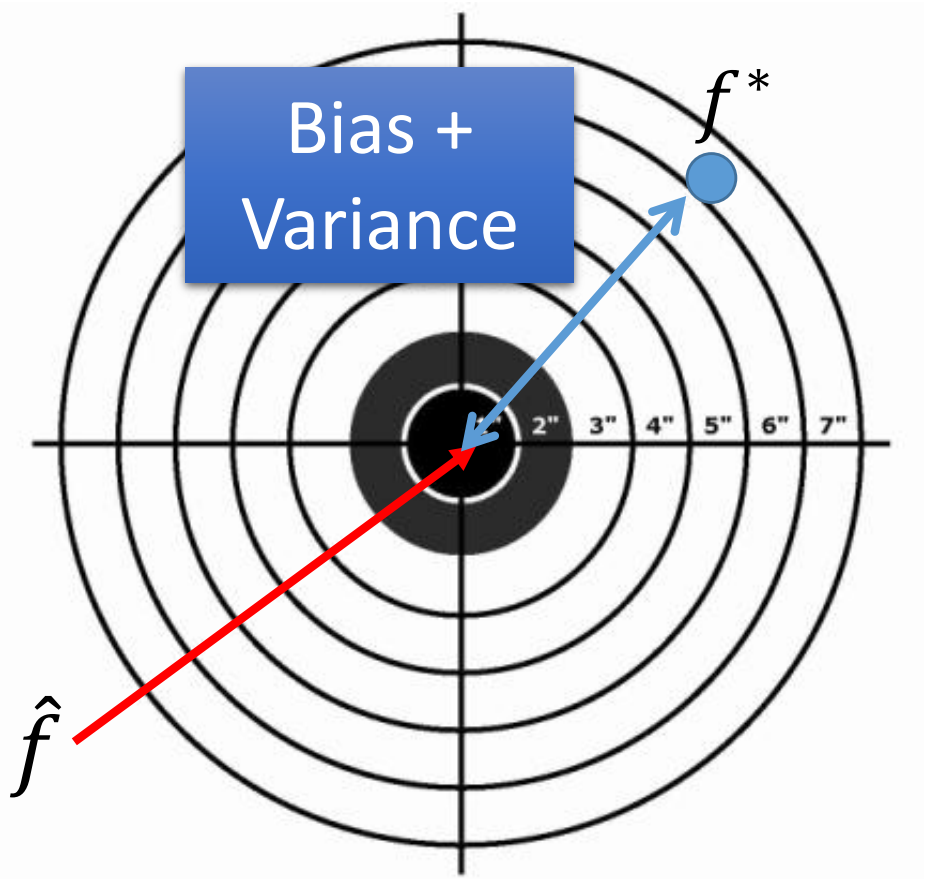
$$\hat{y} = \hat{f}($$ 

$$)$$

Only Niantic knows $\hat{f}$

From training data, we find f^*

f^* is an approximation of $\hat{f}$

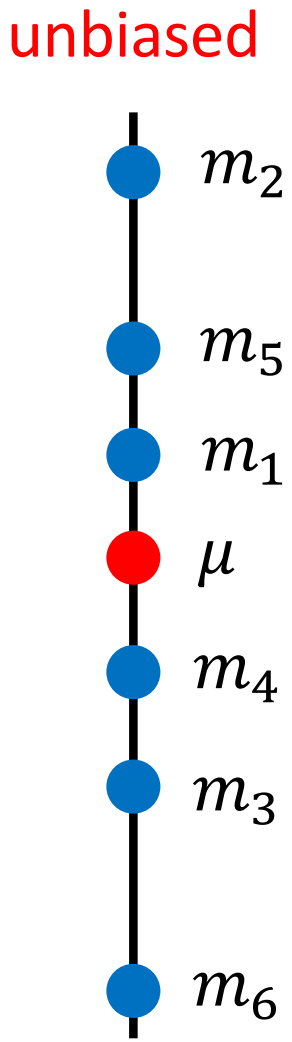


Bias and Variance of Estimator

- Estimate the mean of a variable x
  - assume the mean of x is μ
  - assume the variance of x is σ^2
- Estimator of mean μ
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

$$m = \frac{1}{N}\sum_n x^n \neq \mu$$

$$E[m] = E\left[\frac{1}{N}\sum_n x^n\right] = \frac{1}{N}\sum_n E[x^n] = \mu$$



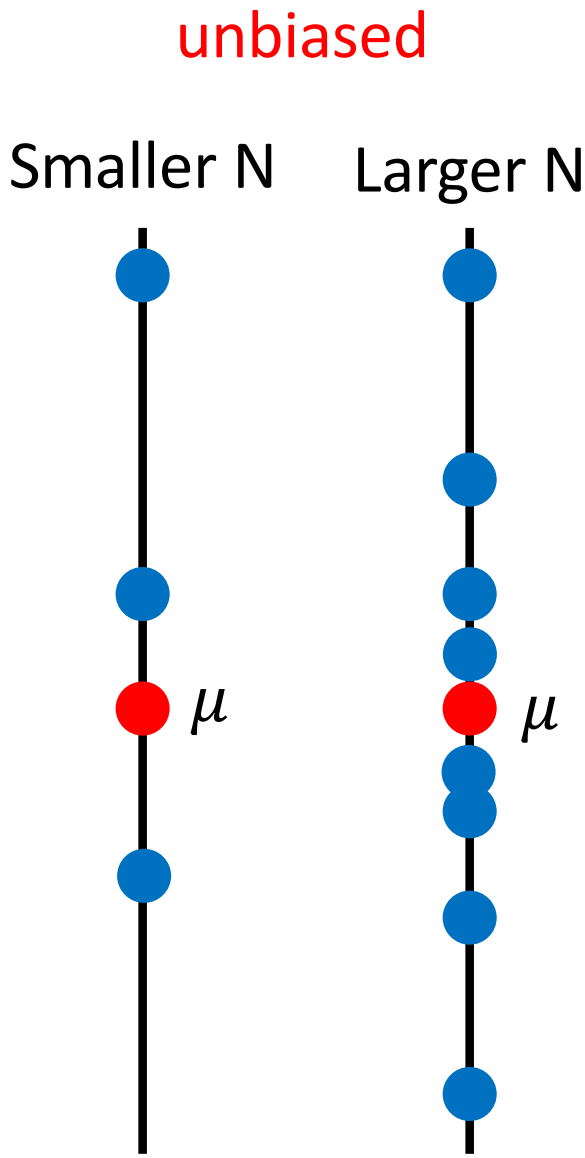
Bias and Variance of Estimator

- Estimate the mean of a variable x
  - assume the mean of x is μ
  - assume the variance of x is σ^2
- Estimator of mean μ
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

$$m = \frac{1}{N}\sum_n x^n \neq \mu$$

$$\mathrm{Var}[m] = \frac{\sigma^2}{N}$$

Variance depends on the number of samples



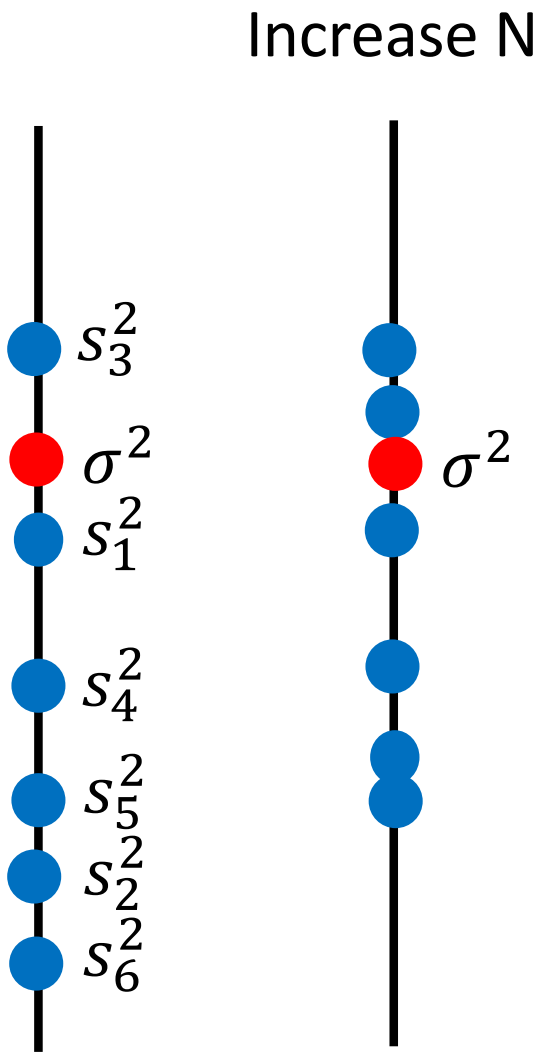
Bias and Variance of Estimator

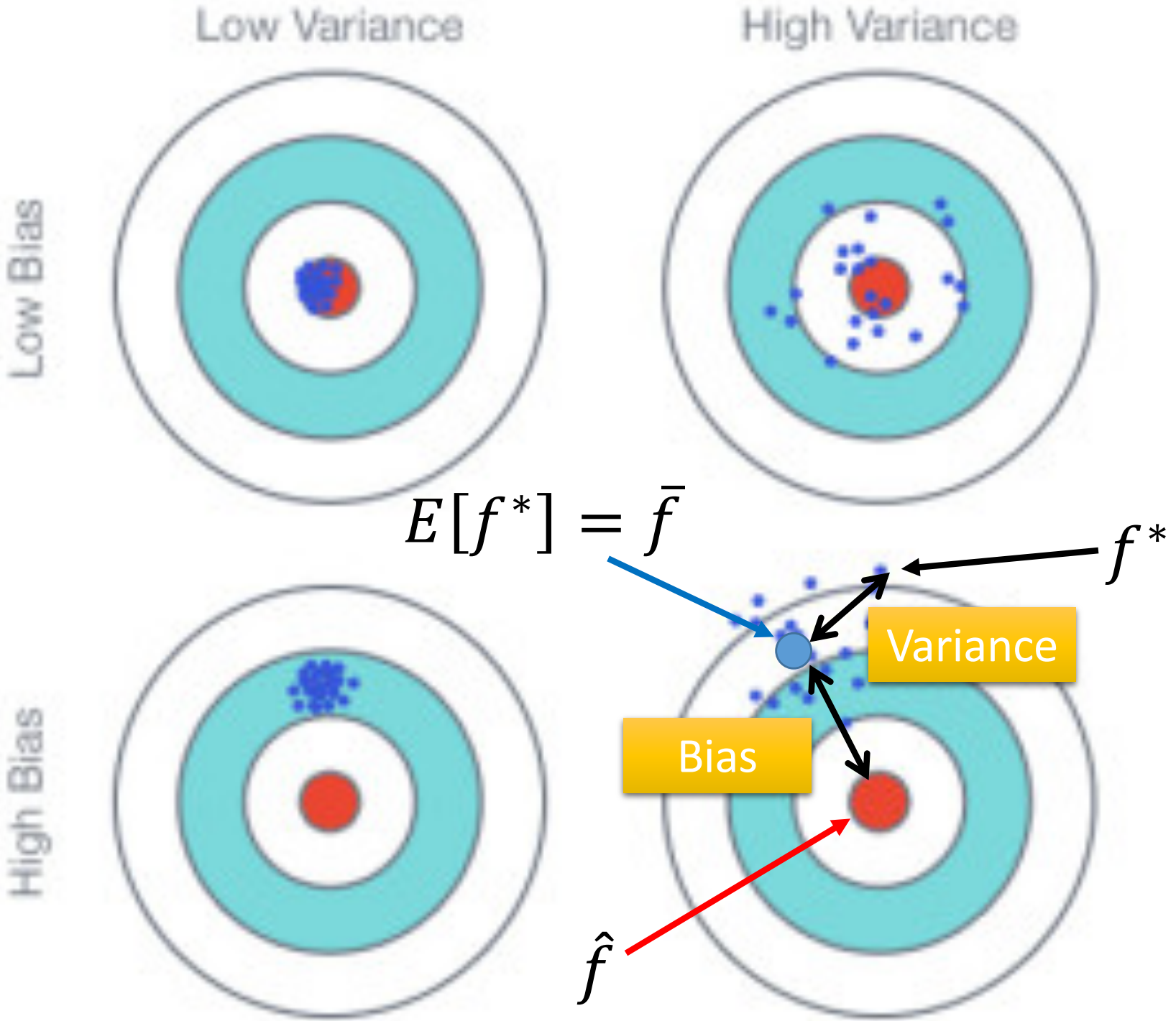
- Estimate the mean of a variable x
  - assume the mean of x is μ
  - assume the variance of x is σ^2
- Estimator of variance σ^2
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

$$m = \frac{1}{N}\sum_n x^n \quad s = \frac{1}{N}\sum_n (x^n - m)^2$$

Biased estimator

$$E[s] = \frac{N-1}{N}\sigma^2 \neq \sigma^2$$

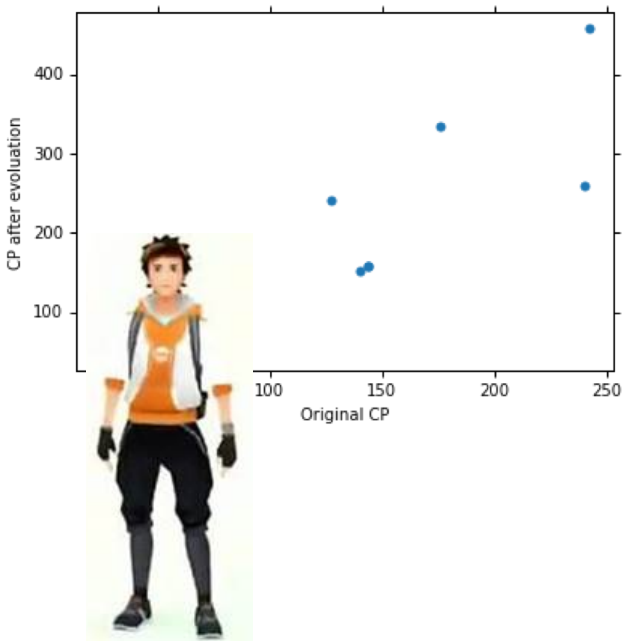



Low Variance
High Variance
Low Bias
High Bias
$E[f^*] = \bar{f}$
f^*
Variance
Bias
$\hat{f}$

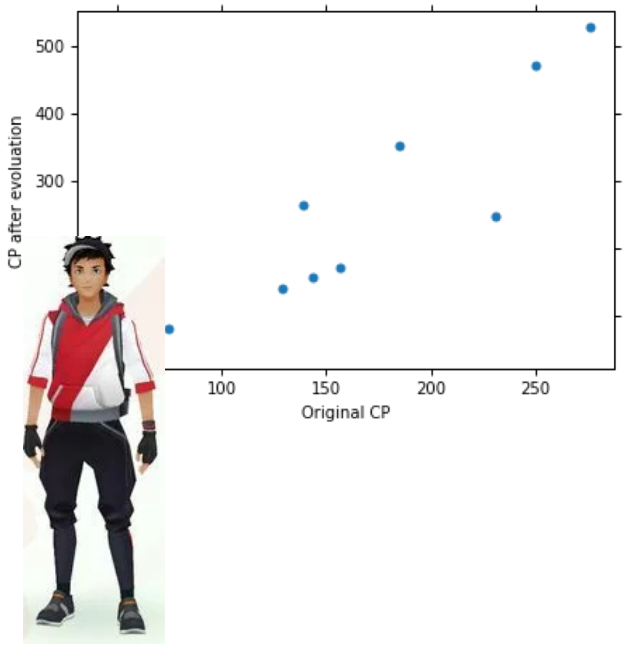
Parallel Universes

- In all the universes, we are collecting (catching) 10 Pokémons as training data to find f^*

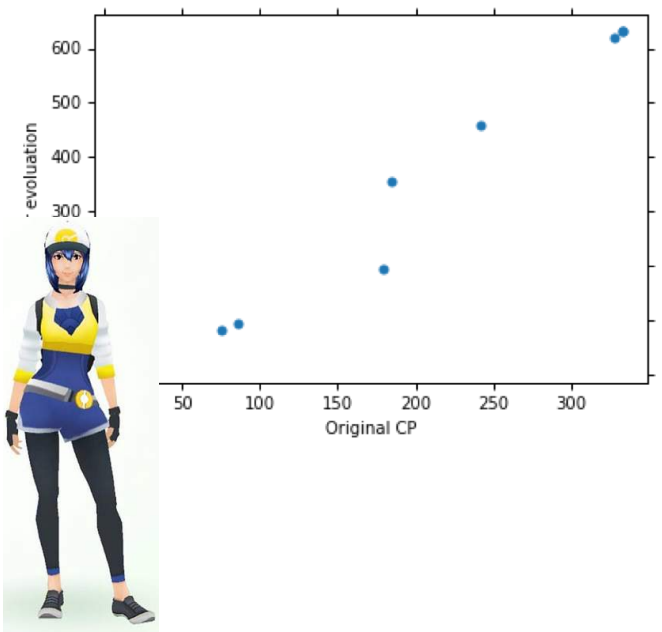
Universe 1



Universe 2



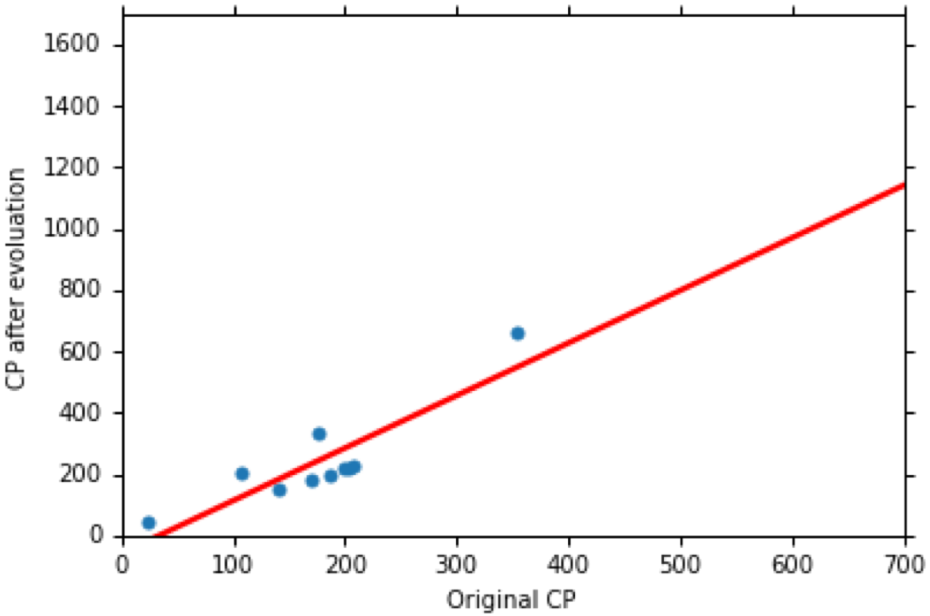
Universe 3



Parallel Universes

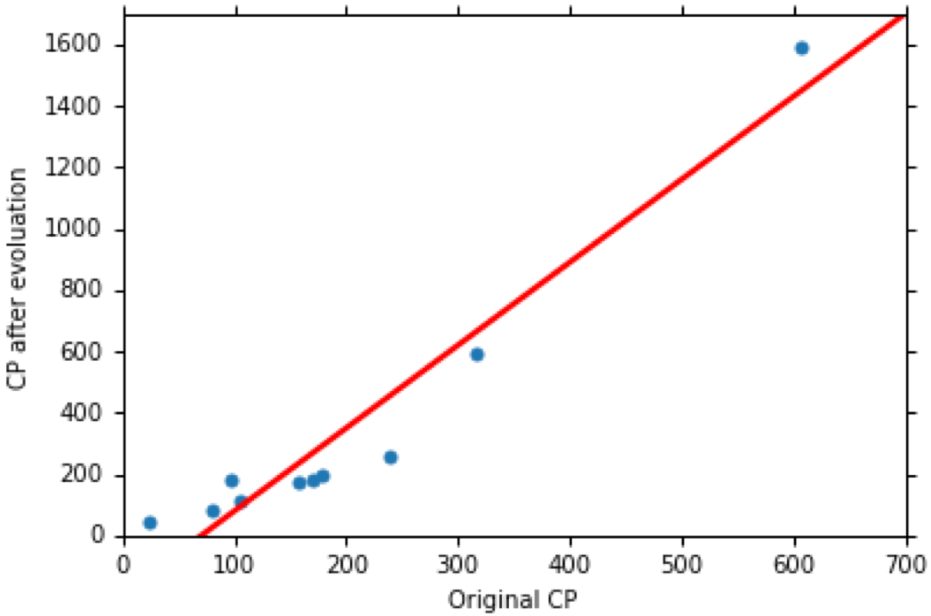
- In different universes, we use the same model, but obtain different f^*

Universe 81



$y = b + w \cdot x_{cp}$

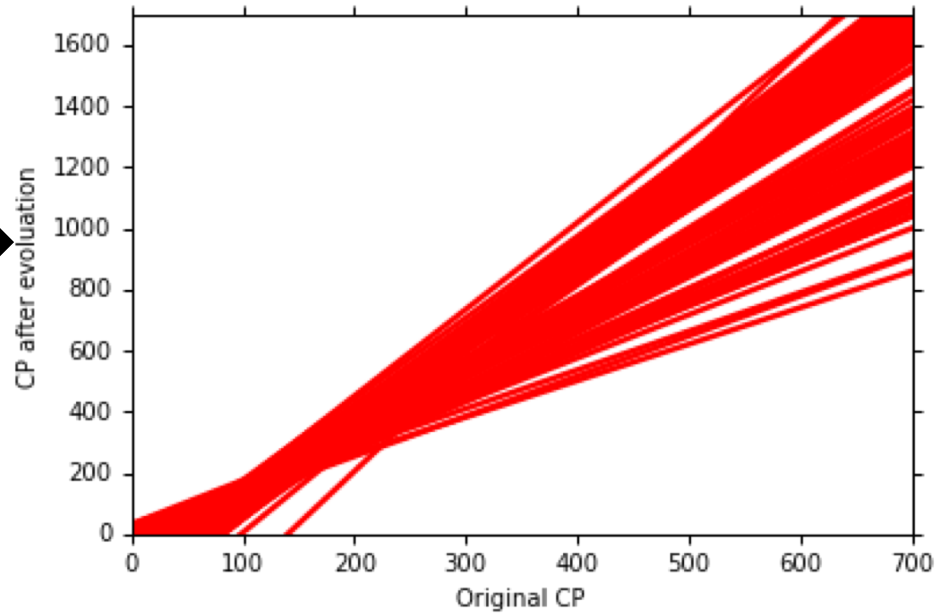
Universe 32

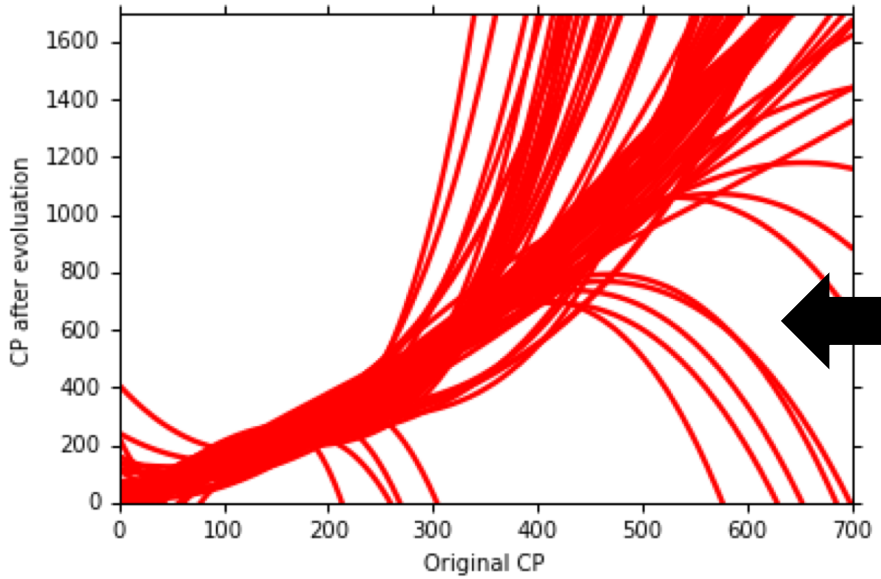


$y = b + w \cdot x_{cp}$

f^* in 100 Universes

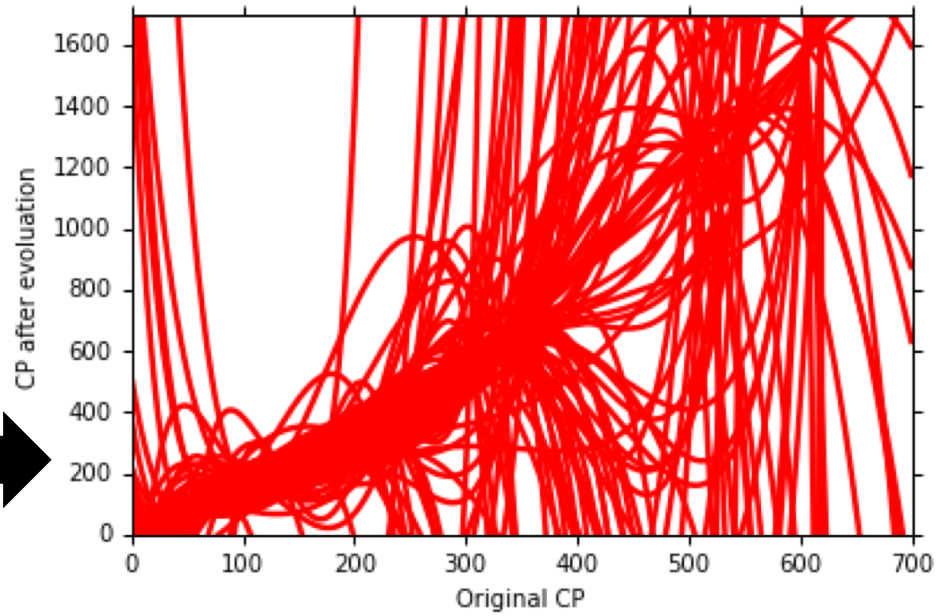
$y = b + w \cdot x_{cp}$





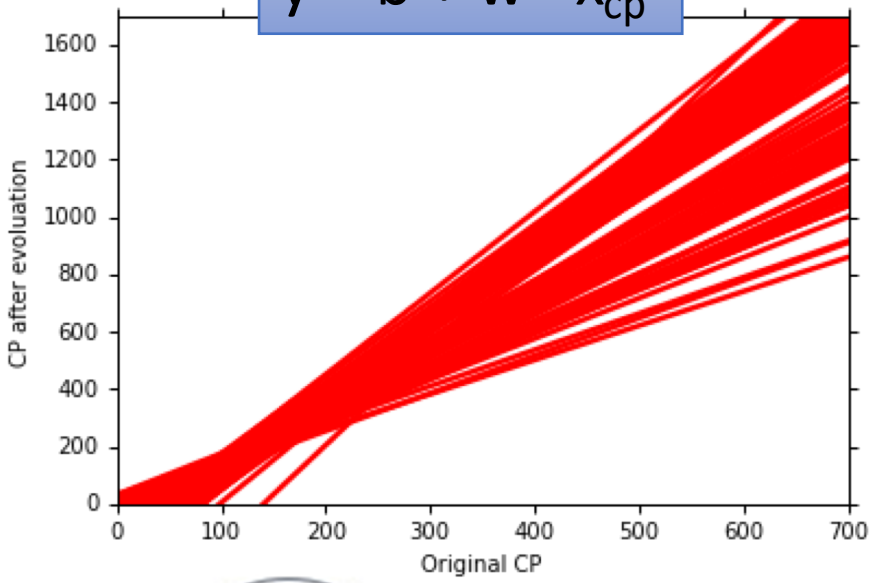
$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3$

$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$

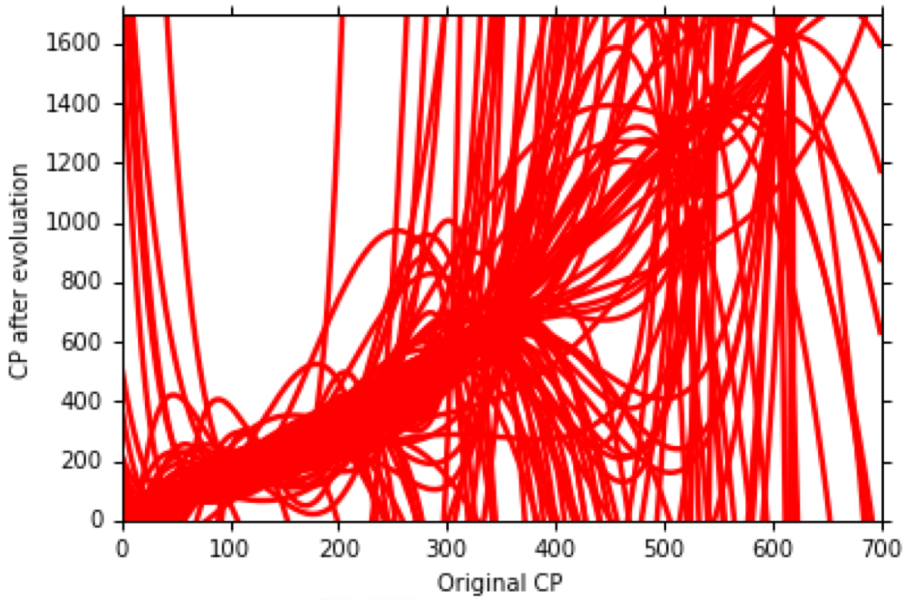


Variance

$y = b + w \cdot x_{cp}$



$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$



Small Variance

Large Variance

Simpler model is less influenced by the sampled data

Consider the extreme case f(x) = 5

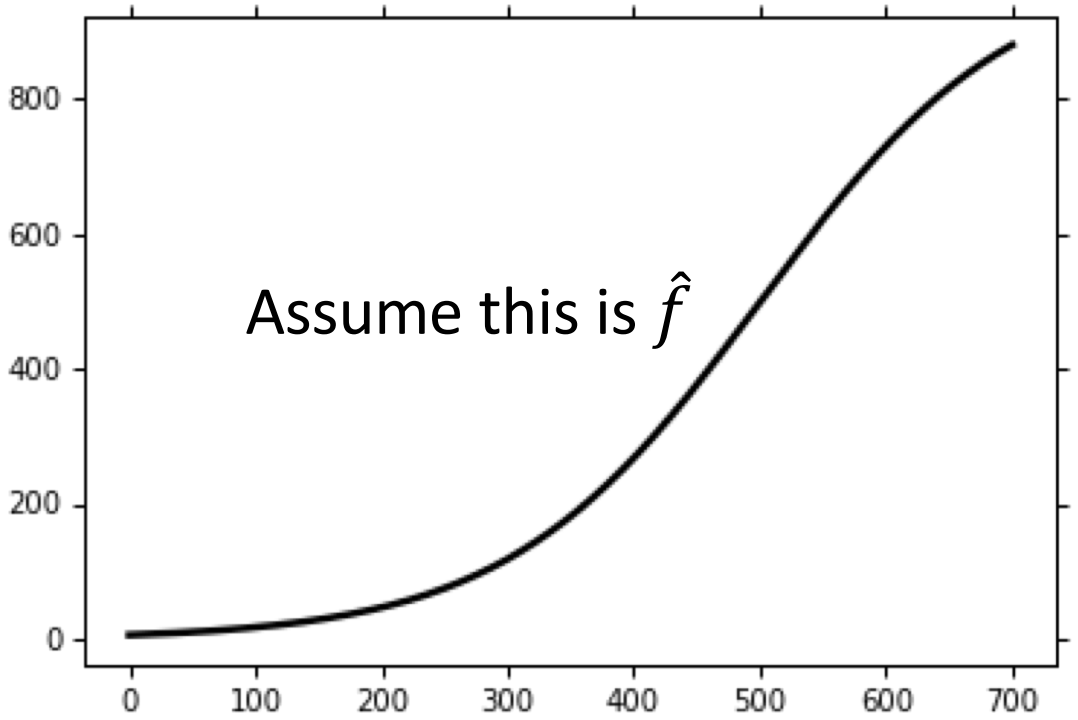
Bias

$$E[f^*] = \bar{f}$$

- Bias: If we average all the f^*, is it close to $\hat{f}$?

Large Bias

Small Bias



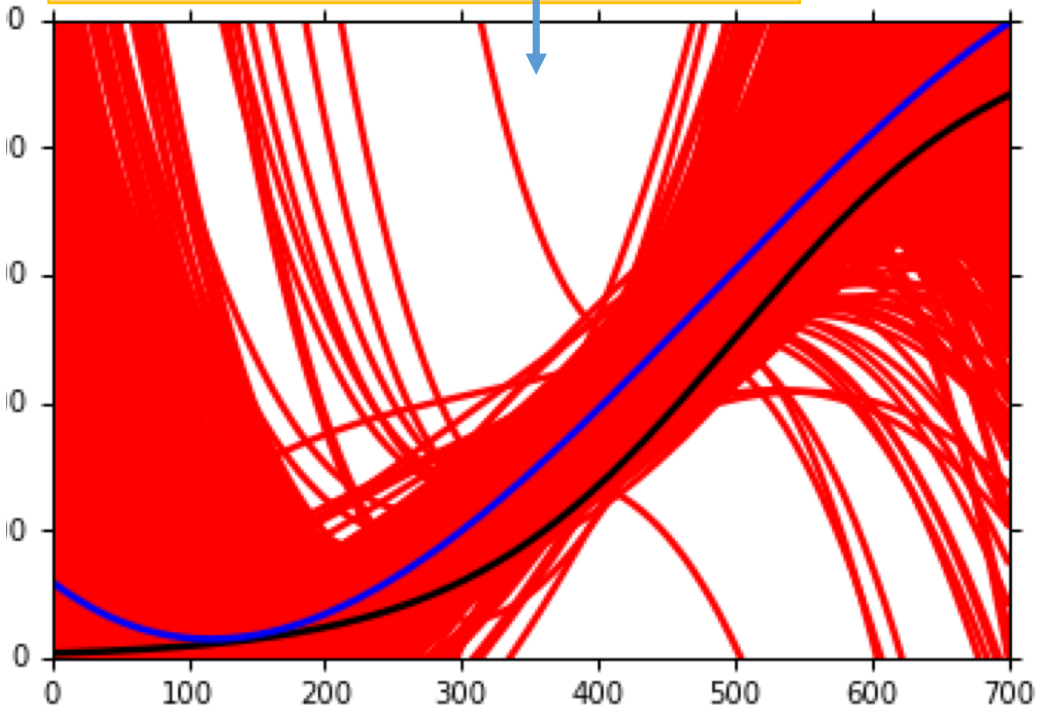
Black curve: the true function $\hat{f}$

Red curves: 5000 f^*

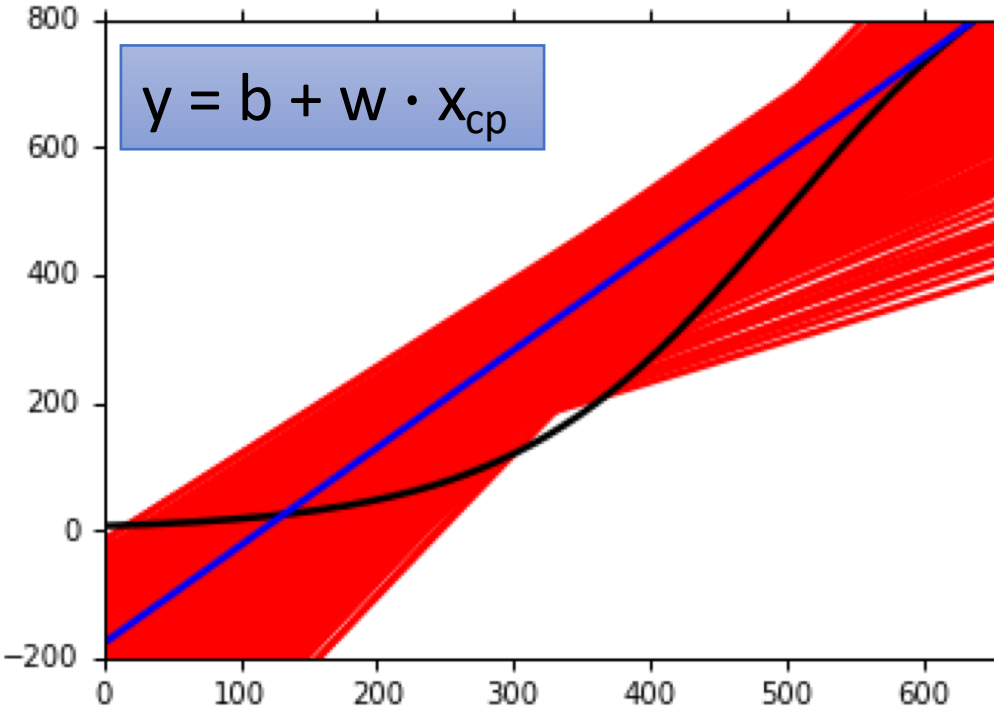
Blue curve: the average of 5000 f^*

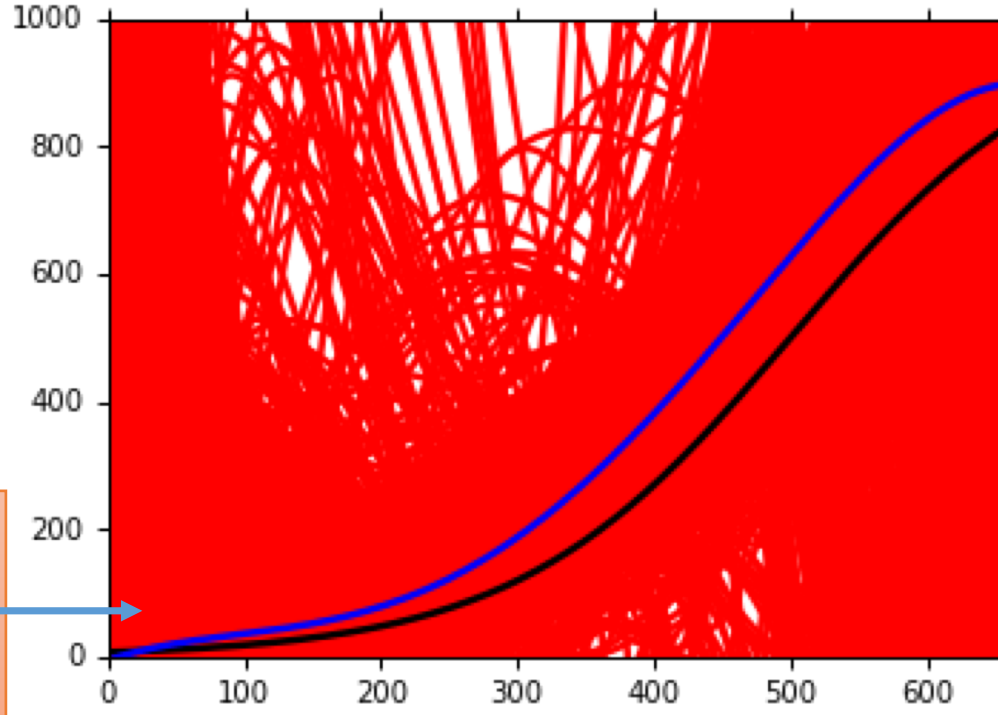
$= \bar{f}$

$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3$



$y = b + w \cdot x_{cp}$



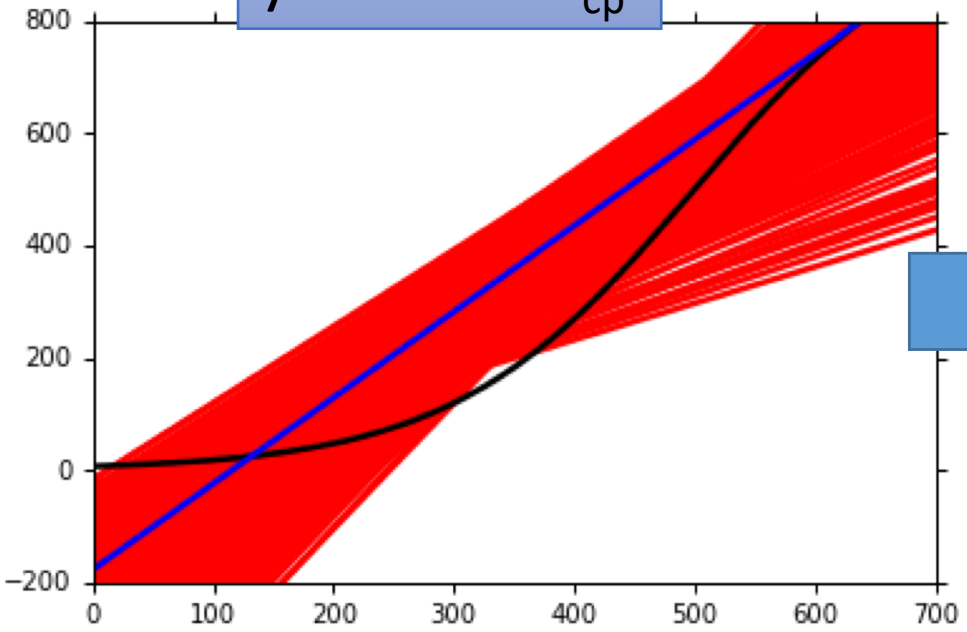


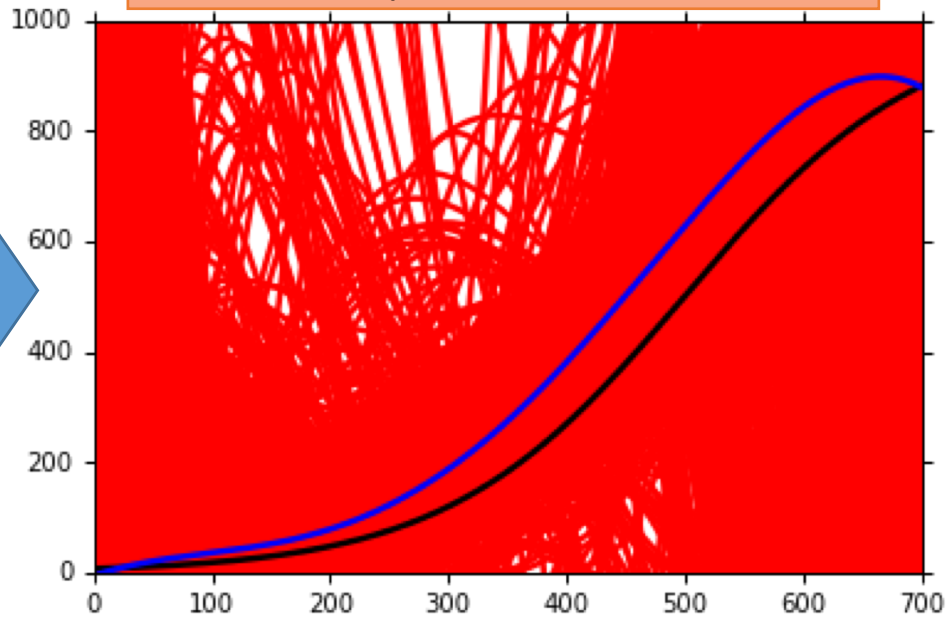
$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$

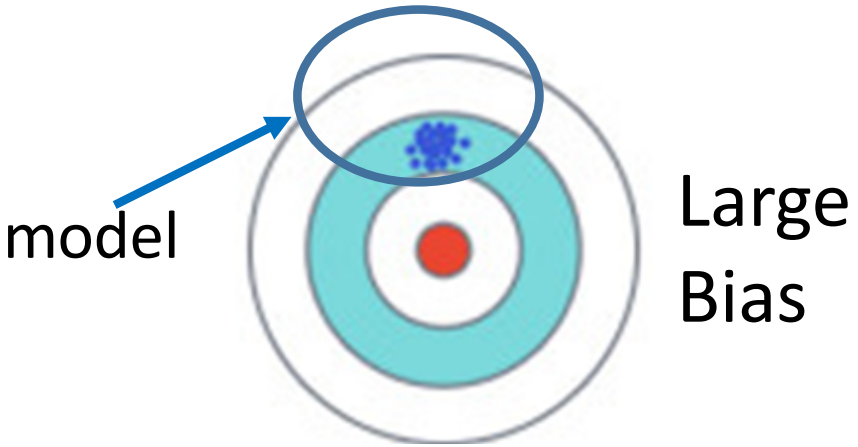
Bias

$y = b + w \cdot x_{cp}$

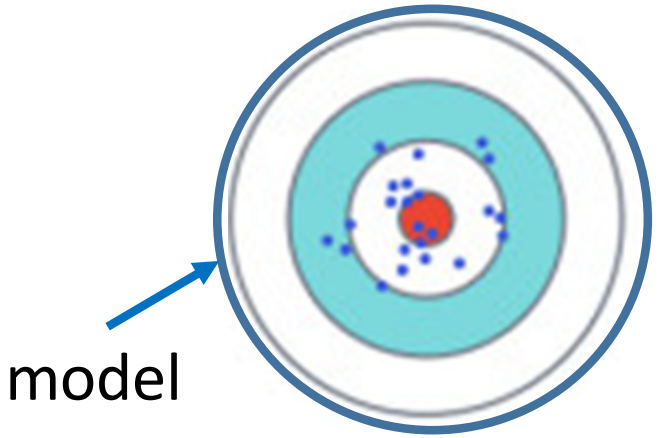
$y = b + w_1 \cdot x_{cp} + w_2 \cdot (x_{cp})^2 + w_3 \cdot (x_{cp})^3 + w_4 \cdot (x_{cp})^4 + w_5 \cdot (x_{cp})^5$





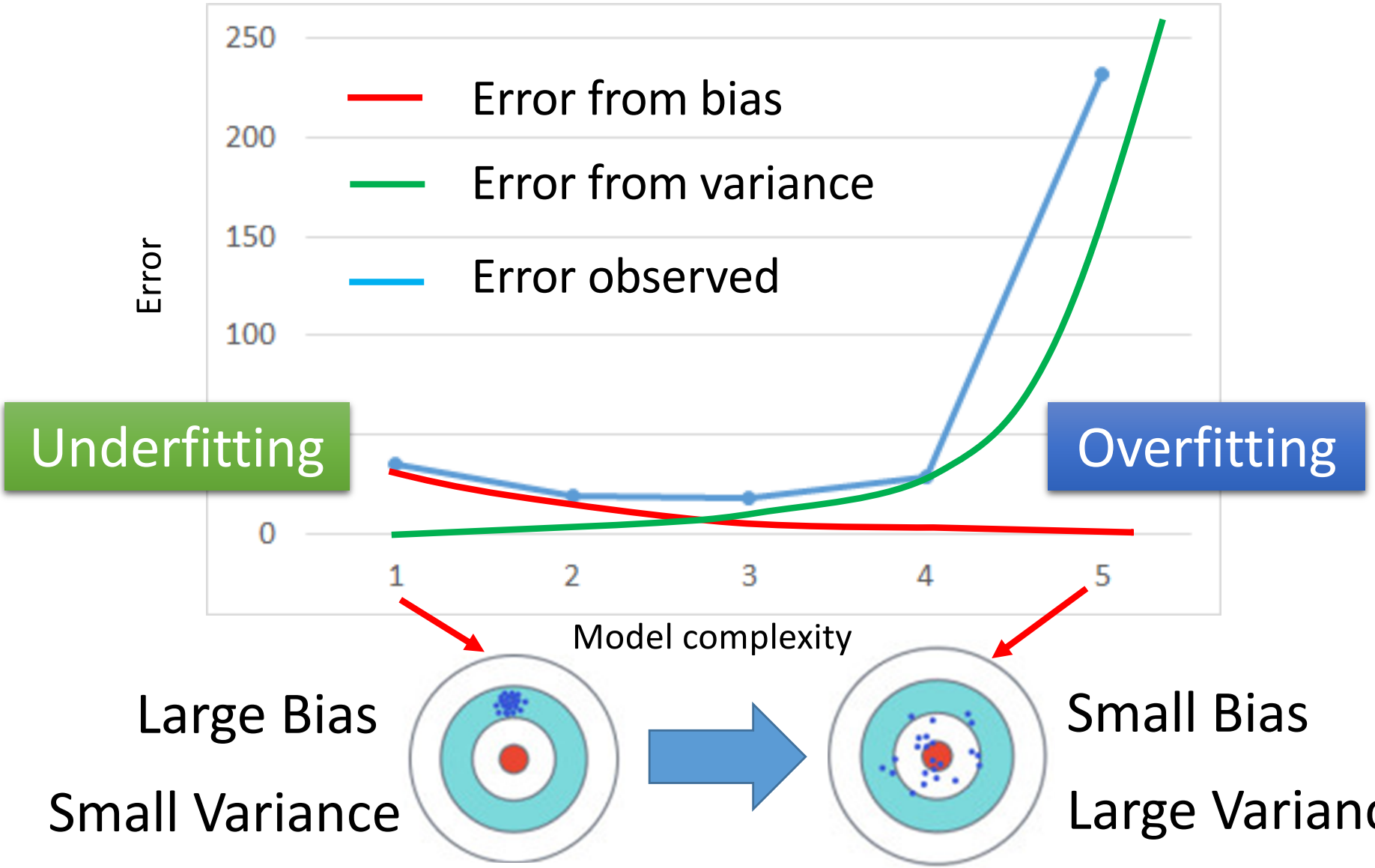


Large Bias



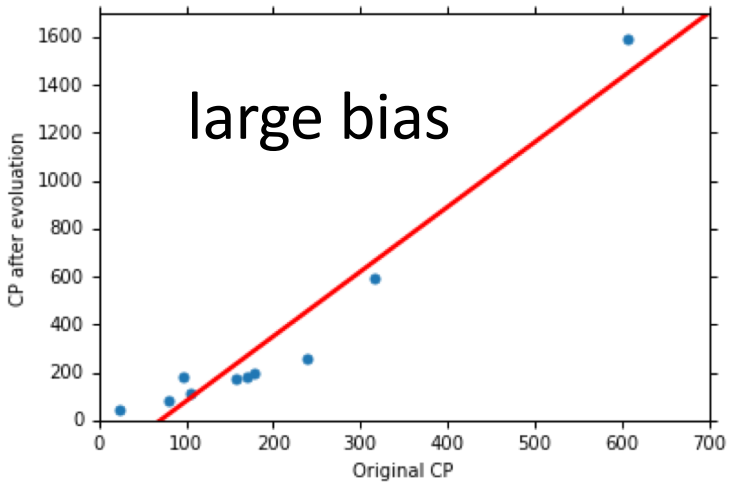
Small Bias

Bias v.s. Variance



What to do with large bias?

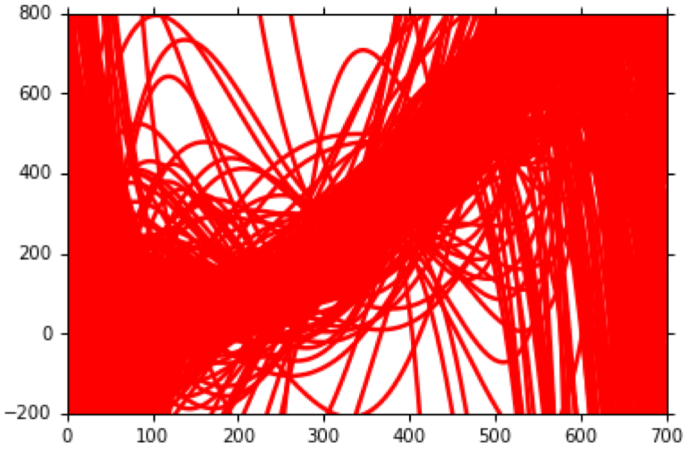
- Diagnosis:
  - If your model cannot even fit the training examples, then you have large bias Underfitting
  - If you can fit the training data, but large error on testing data, then you probably have large variance Overfitting
- For bias, redesign your model:
  - Add more features as input
  - A more complex model



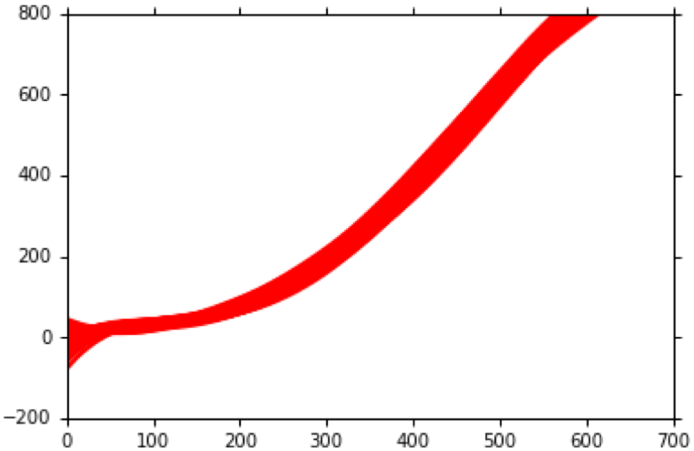
What to do with large variance?

- More data

Very effective, but not always practical

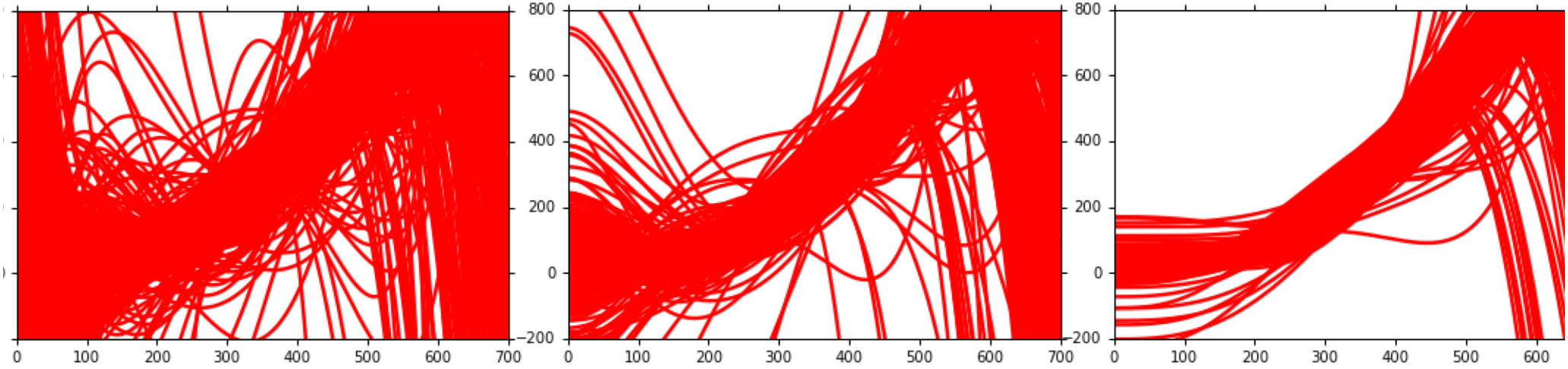


10 examples



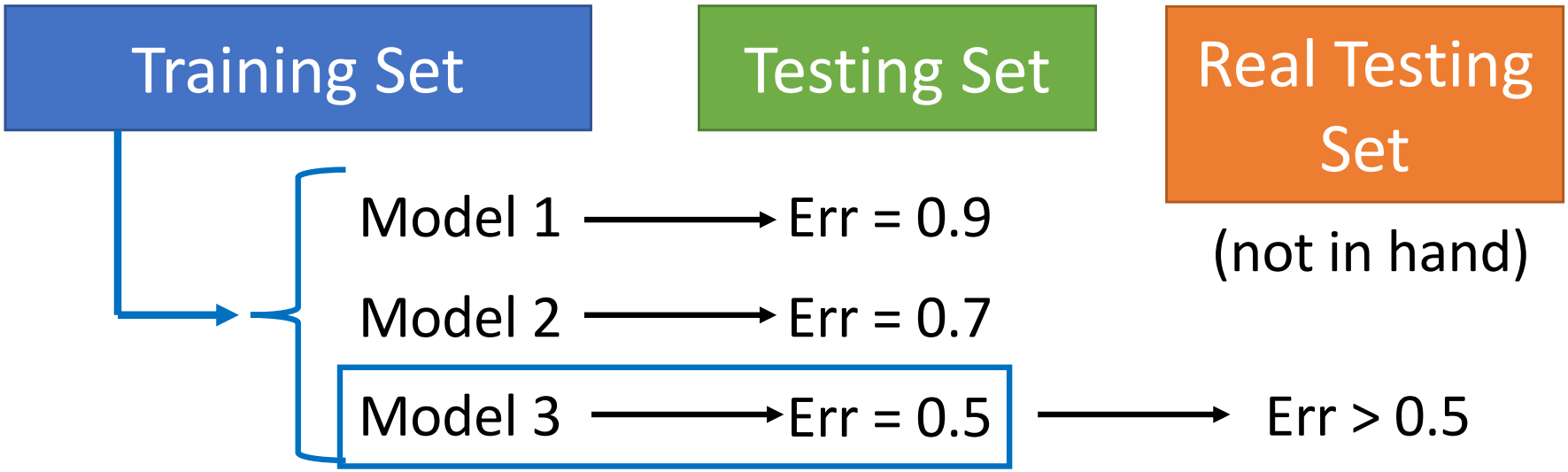
100 examples

- Regularization ➡ May increase bias

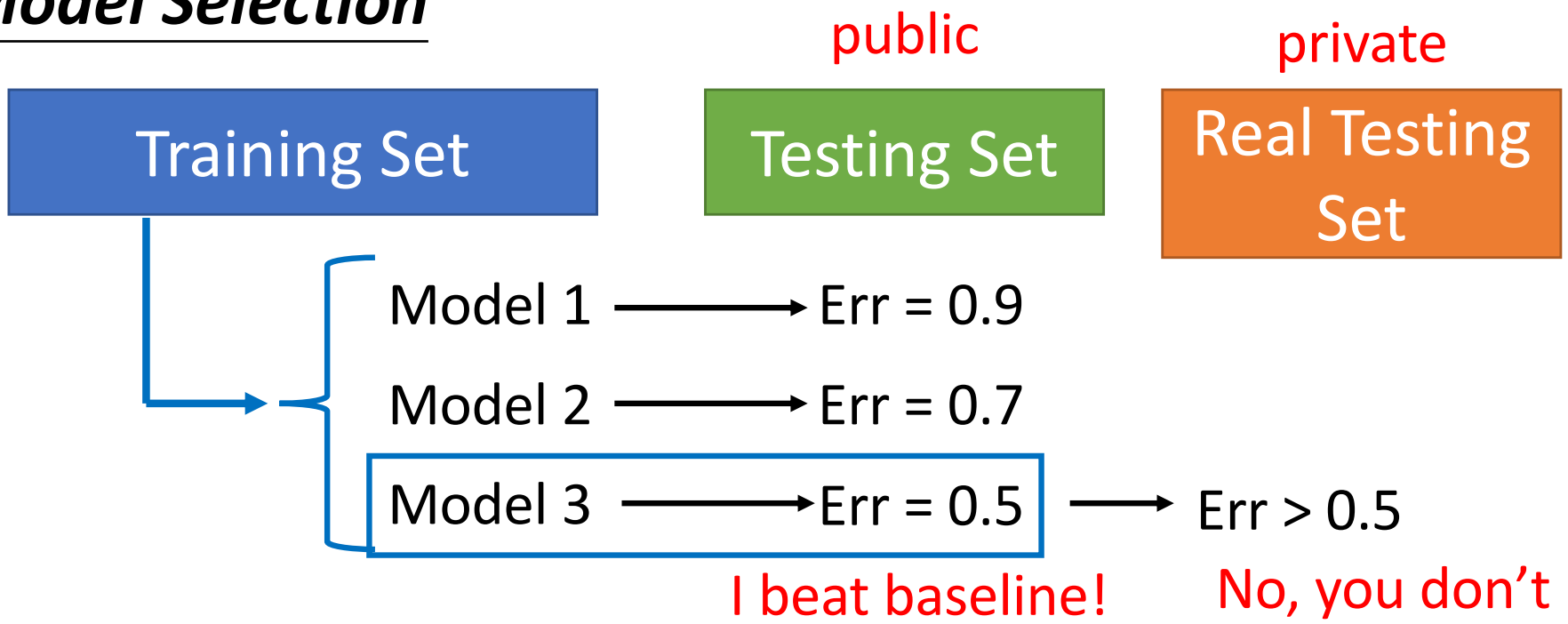


Model Selection

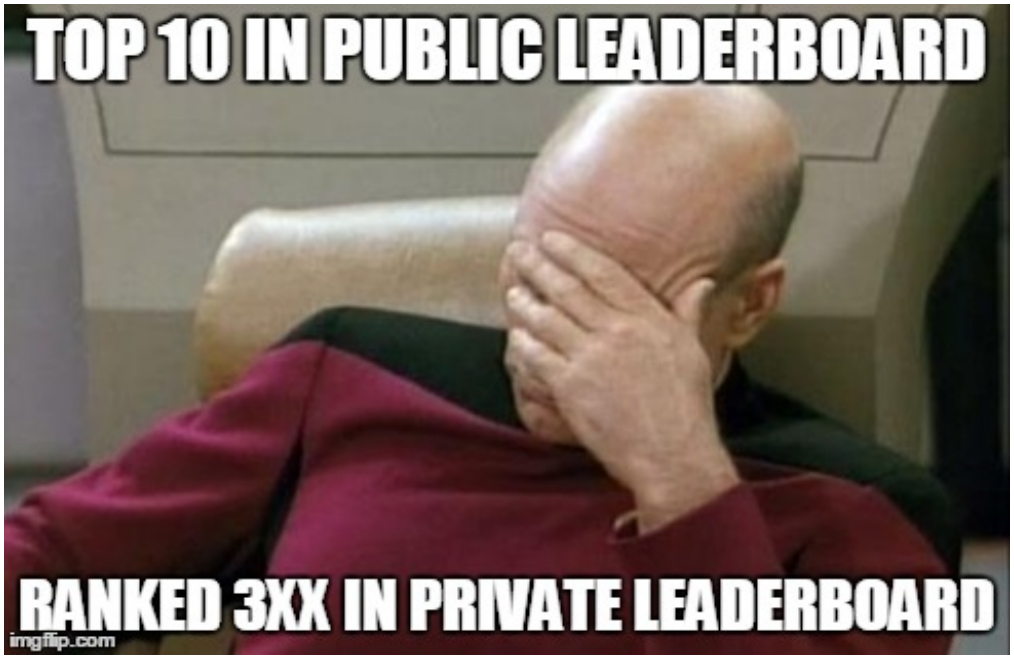
- There is usually a trade-off between bias and variance.
- Select a model that balances two kinds of error to minimize total error
- What you should NOT do:



Model Selection

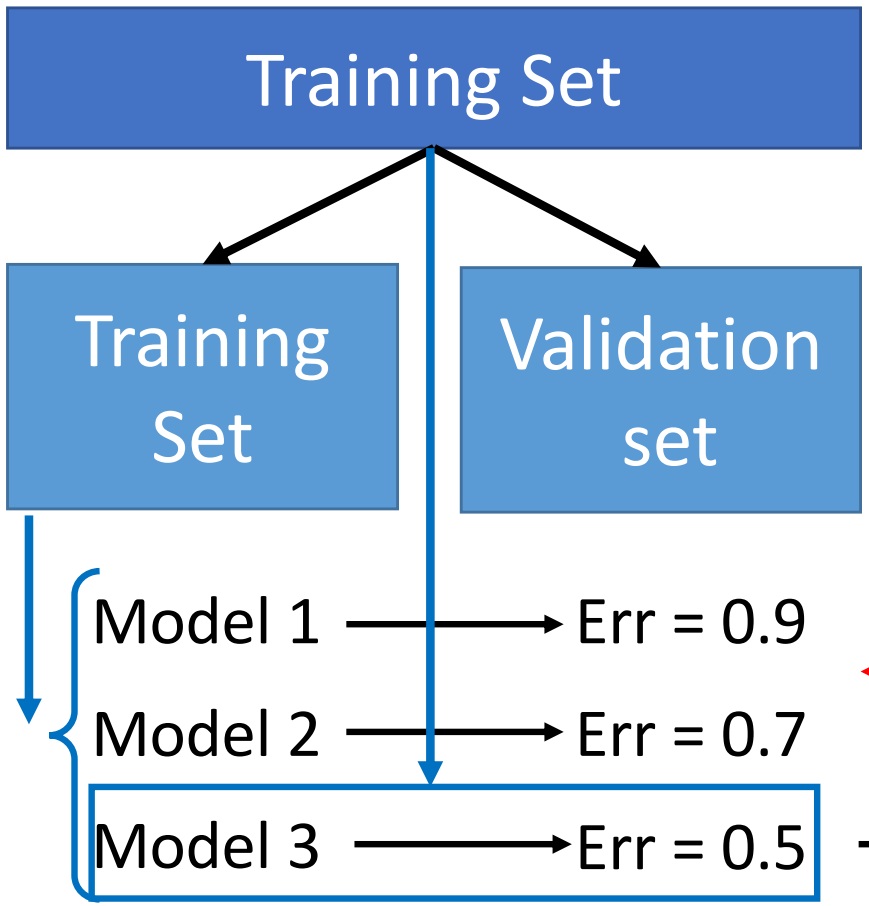


What happened?



http://www.chioka.in/how-to-select-your-final-models-in-a-kaggle-competitio/

Cross Validation



public

Testing Set

private

Real Testing Set

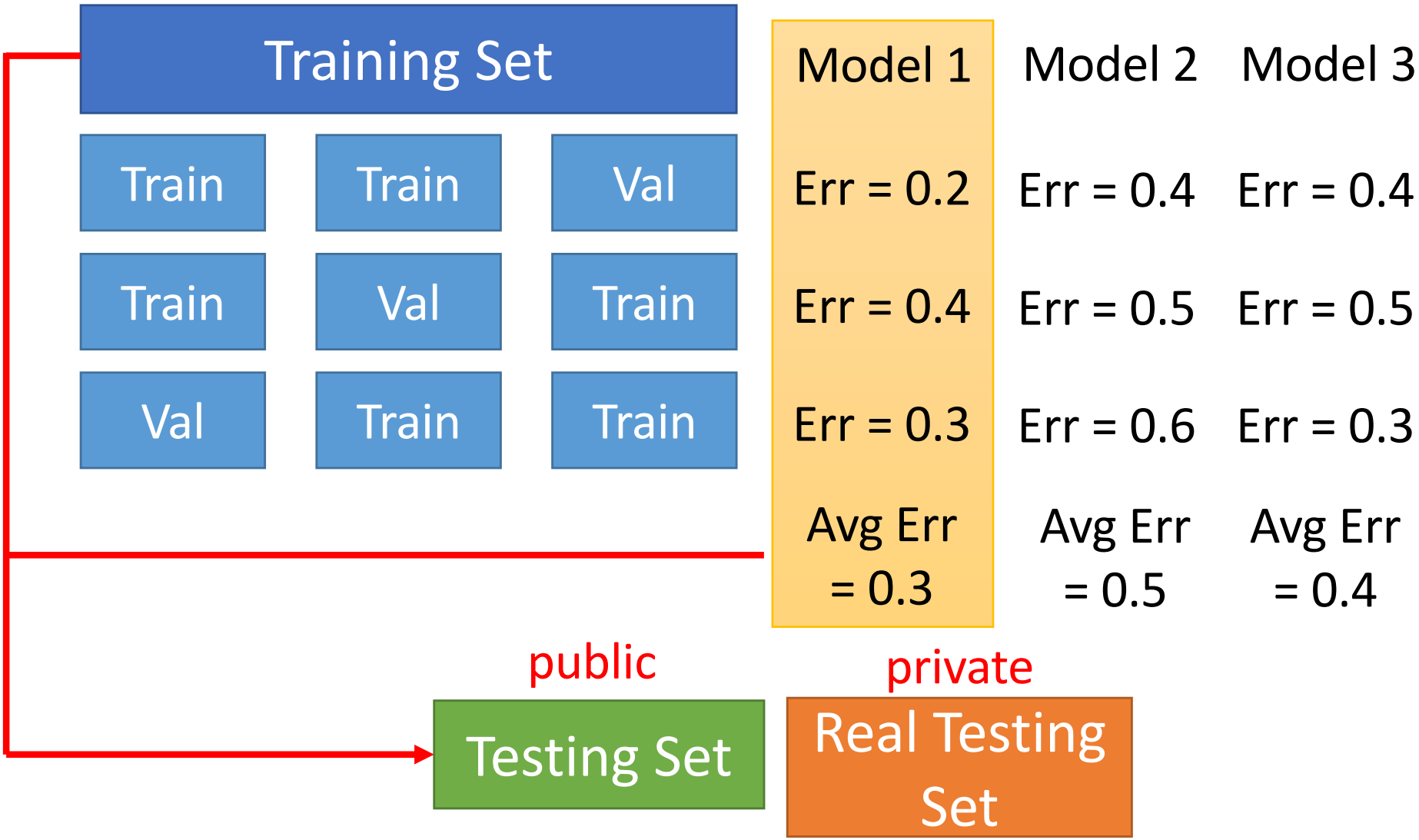
Using the results of public testing data to tune your model

You are making public set better than private set.

Not recommend

Err > 0.5 → Err > 0.5

N-fold Cross Validation



Questions?