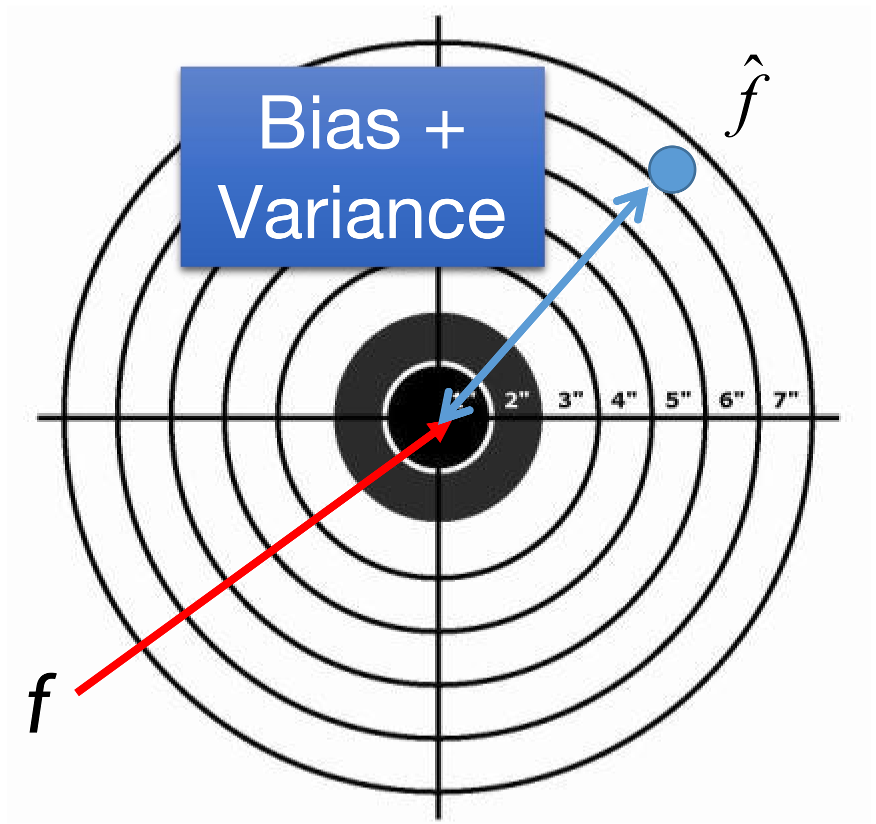
Where does the error come from?

Summary


error
250
200
150
100
50
0
Error from bias
Error from variance
Error on testing data
1
2
3
4
5
model
Large Bias
Small Variance
Small Bias
Large Variance

Estimator

$$y = f(x)$$



From training data,
we find $\hat{f}$

$\hat{f}$ is an estimator of f

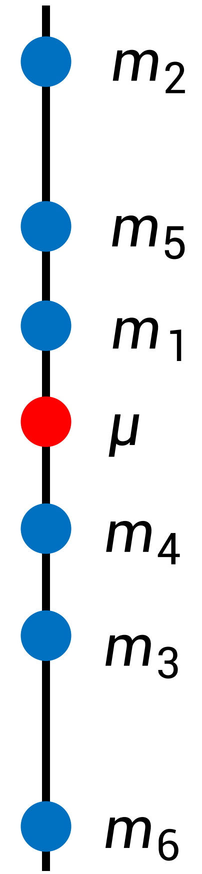
Bias and Variance of Estimator

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

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

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

unbiased



Bias and Variance of Estimator

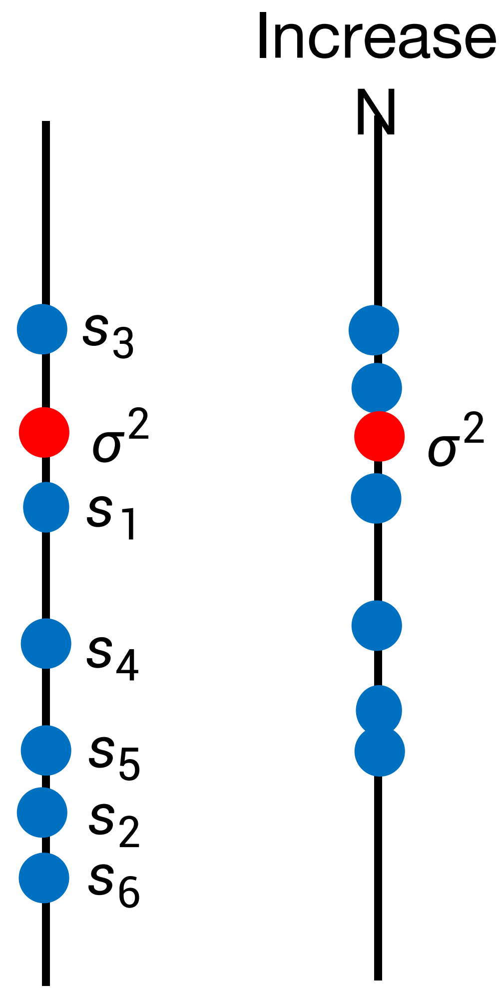
- Estimator of variance σ^2
  - Sample N points: $\{x^1, x^2, \ldots, x^N\}$

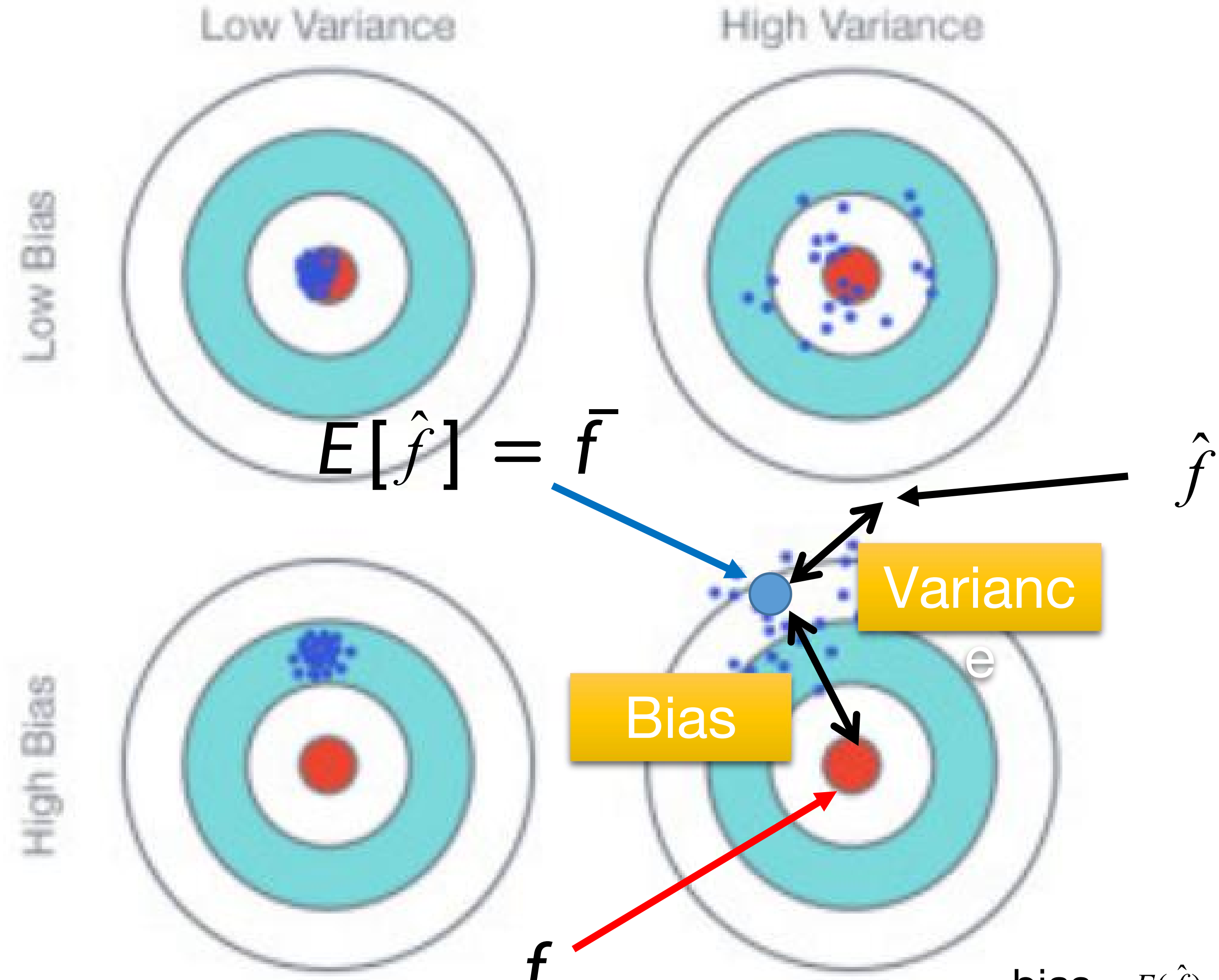
$$m = \frac{1}{N}\sum_n x^n \qquad s = \frac{1}{N}\sum_n (x^n - m)^2$$

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

Biased estimator

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



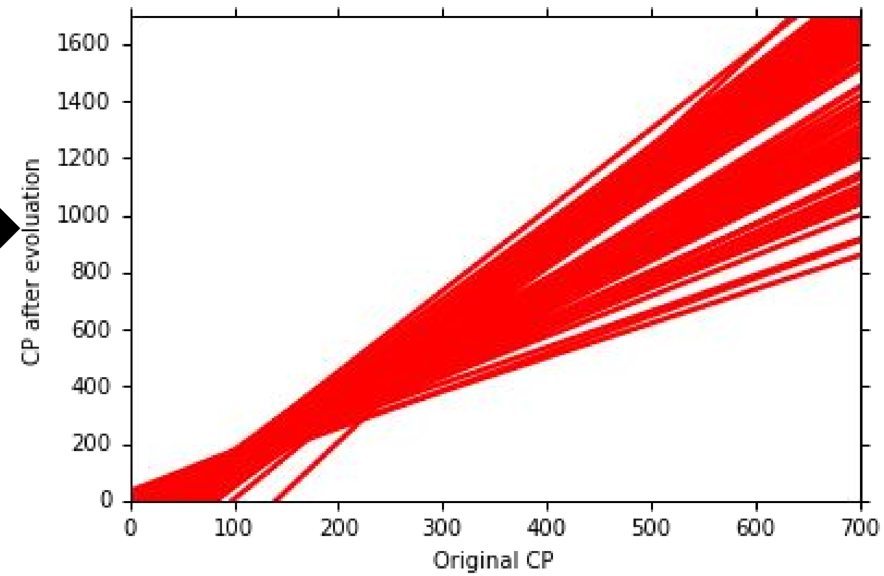


bias $= E(\hat{f}) - f$

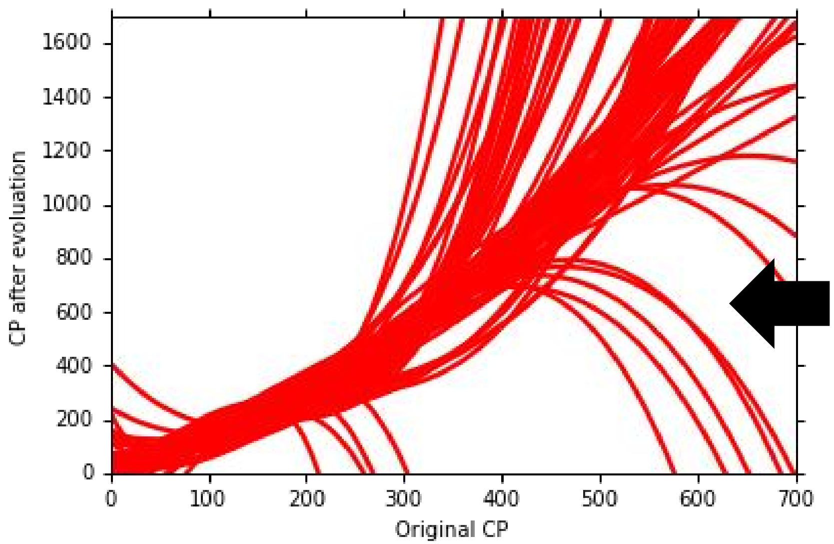
variance $= E(E(\hat{f}) - \hat{f})^2$

$\hat{f}$ in 100 samples

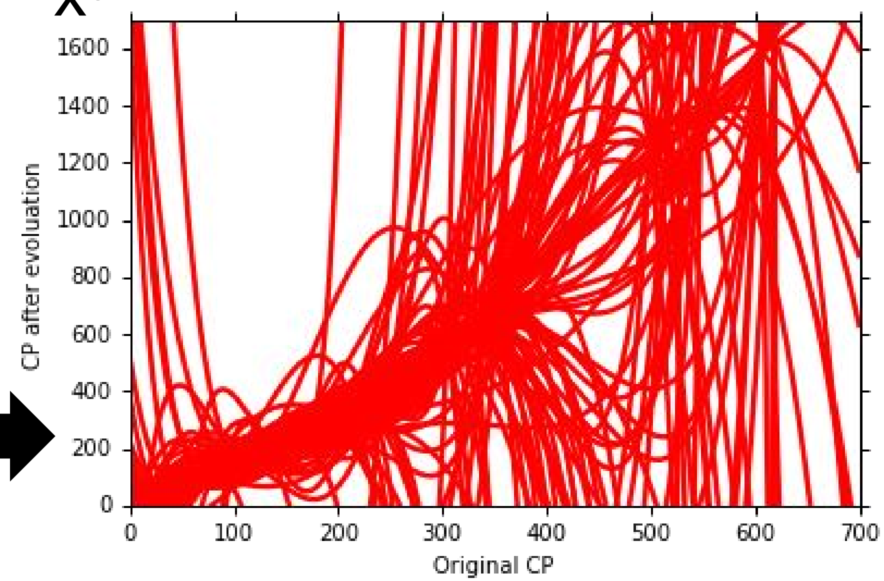
$y = b + w \cdot x$



$y = b + w_1 \cdot x + w_2 \cdot x^2 + w_3 \cdot x^3$

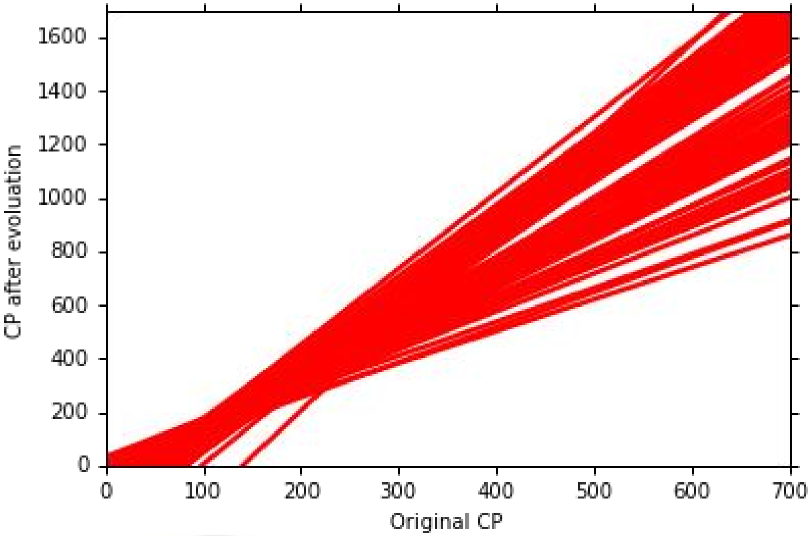


$y = b + w_1 \cdot x + w_2 \cdot x^2 + w_3 \cdot x^3 + w_4 \cdot x^4 + w_5 \cdot x^5$

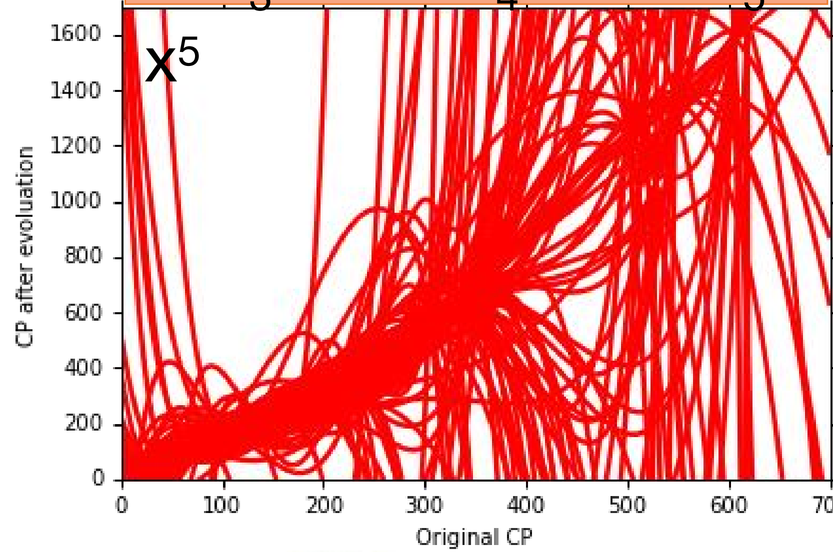


Variance

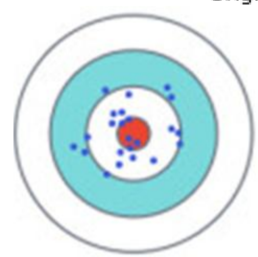
$y = b + w \cdot x$



$y = b + w_1 \cdot x + w_2 \cdot x^2 + w_3 \cdot x^3 + w_4 \cdot x^4 + w_5 \cdot x^5$



Small Variance



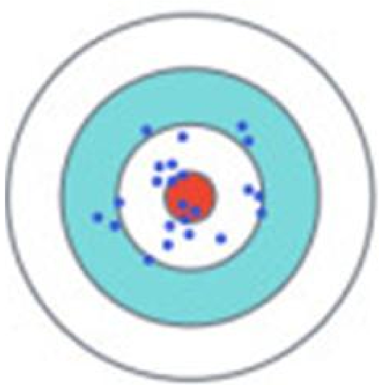
Large Variance

Simpler model is less influenced by the sampled data

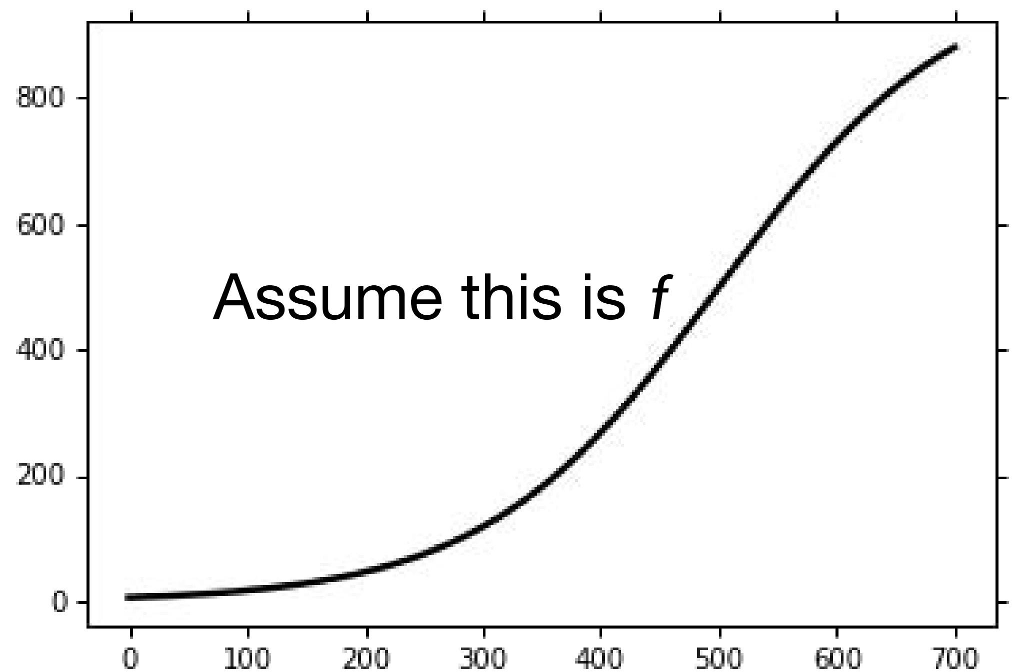
Bias

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

Larg
e
Bias



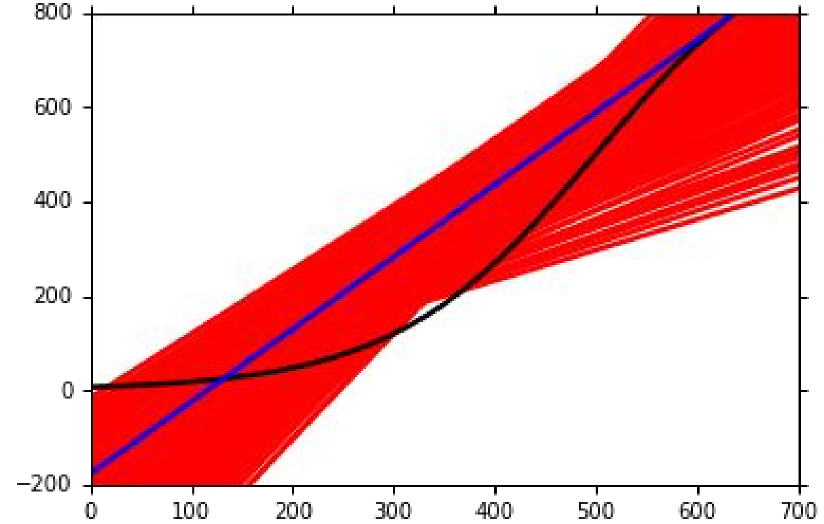
Small
Bias

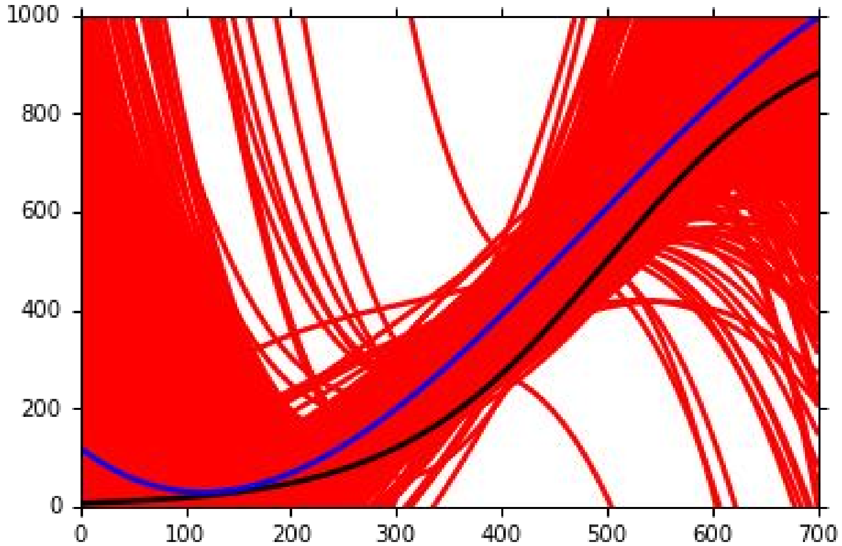


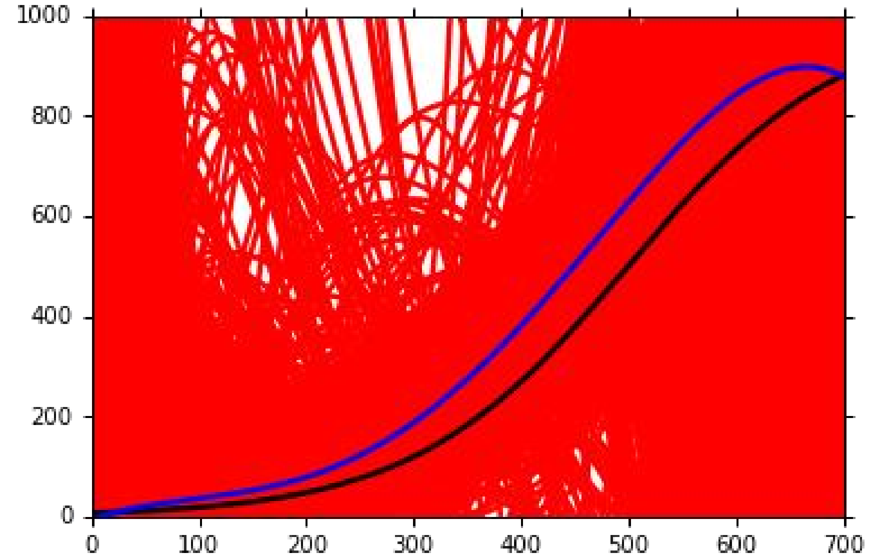
Black curve: the true function f

Red curves: 5000 $\hat{f}$

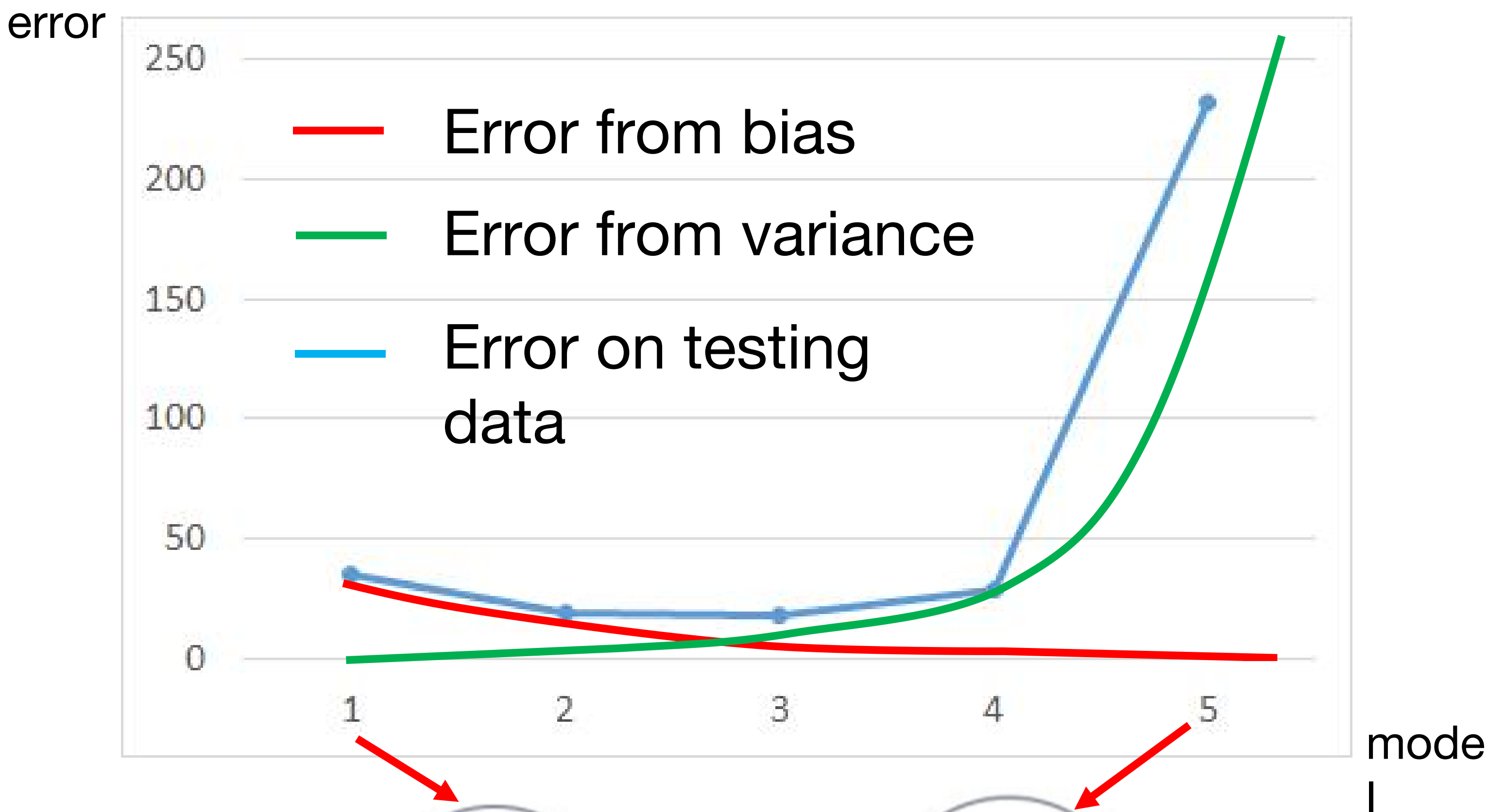
Blue curve: the average of 5000 $\hat{f} = \bar{f}$





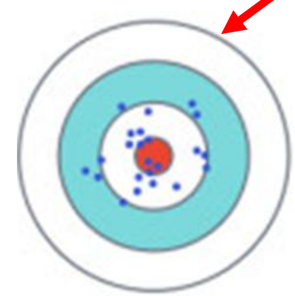


Bias vs. Variance



Large Bias

Small Variance

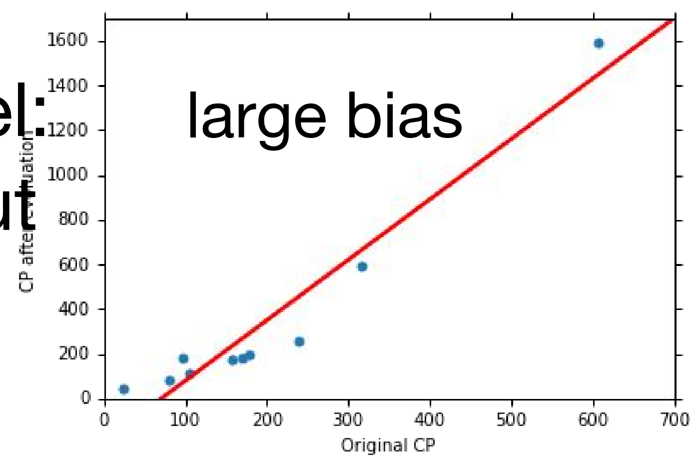


Small Bias

Large Variance

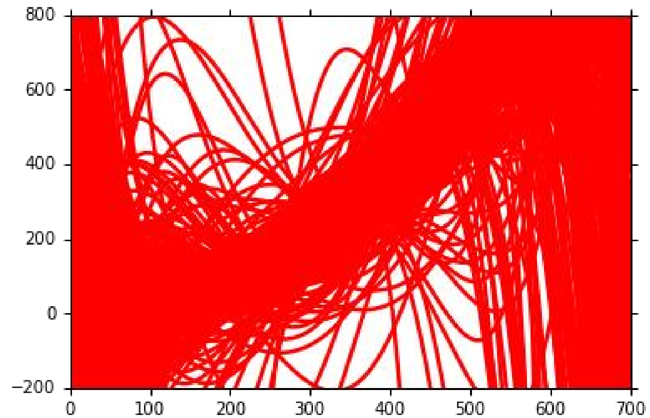
What to do with large bias?

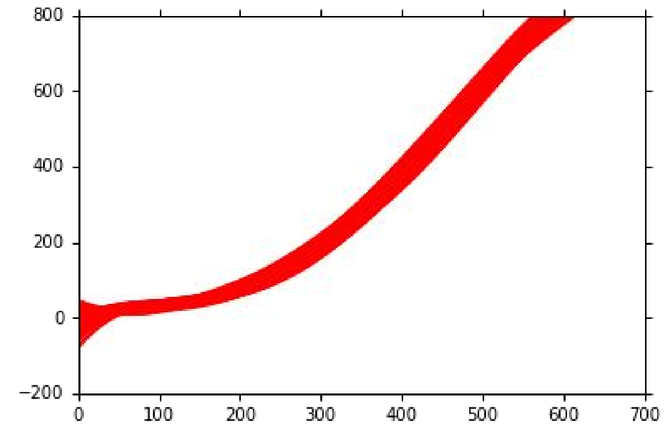
- Diagnosis:
  - If your model cannot even fit the training examples, then you have large bias
  - If you can fit the training data, but large error on testing data, then you probably have large variance
- For bias, redesign your model:
  - Add more features as input
  - A more complex model



What to do with large variance?

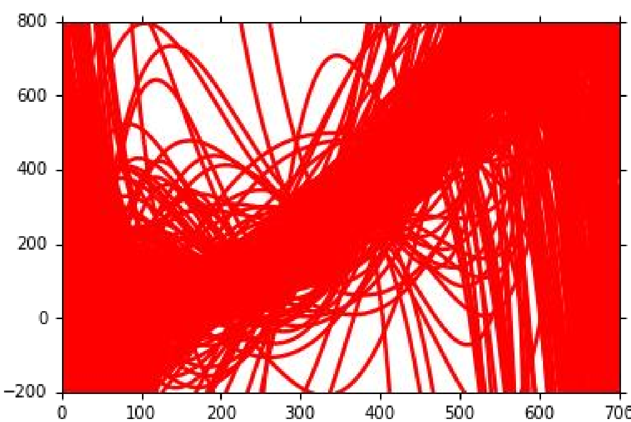
- More data

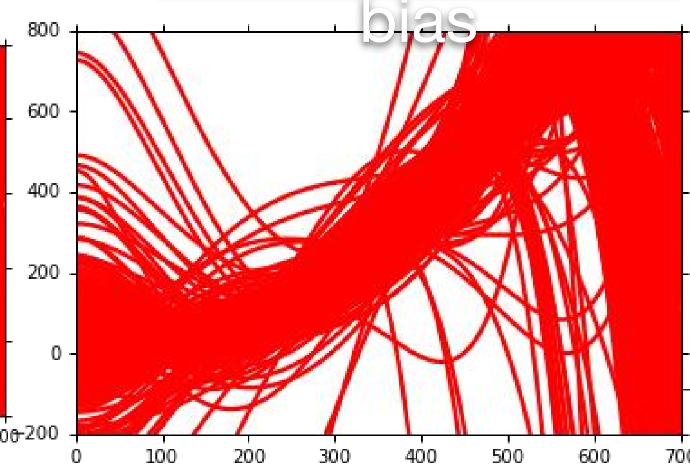


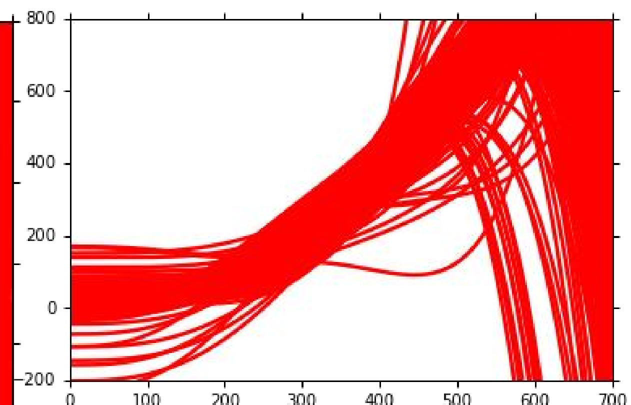


- Regularization

May increase bias







Ex. of Regularization

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

$$L = \sum_n \left(y^n - \left(b + \sum w_i x_i^n\right)\right)^2$$

The functions with smaller w_i are better

$$\boxed{+\lambda \sum (w_i)^2}$$

- Smooth functions are preferred
- If some noises corrupt input when testing, a smoother function has less influence.

Regularization

smoother

λ	Training	Testing
0	1.9	102.3
1	2.3	68.7
10	3.5	25.7
100	4.1	11.1
1000	5.6	12.8
10000	6.3	18.7
100000	8.5	26.8

➢ Training error: smaller λ, less the training error