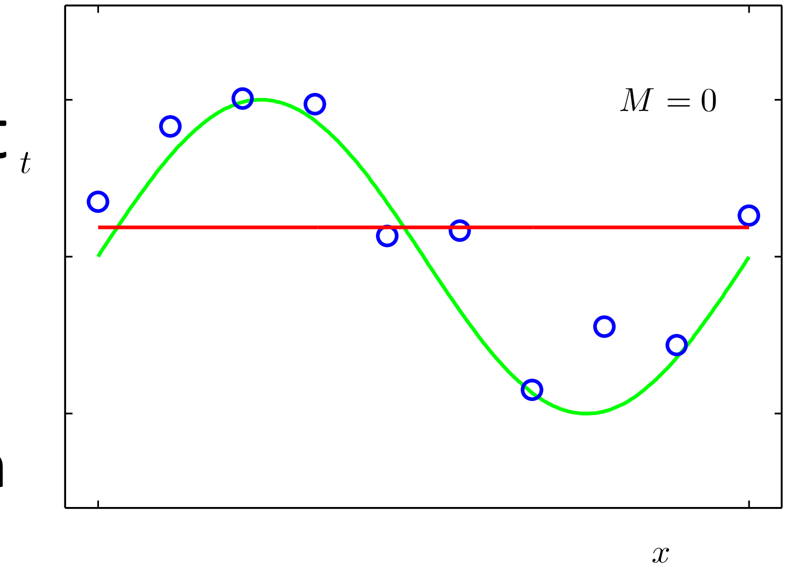
CSE 446
Learning Theory

Administrative

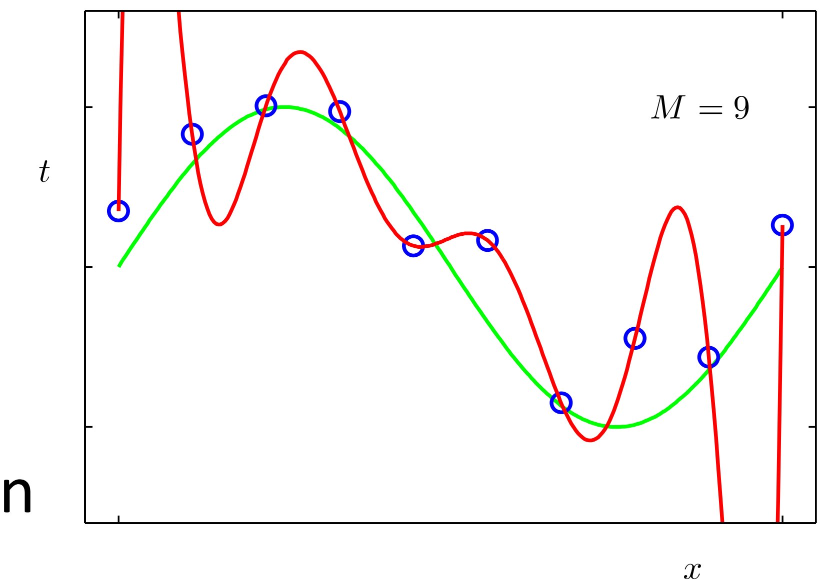
- Quiz section next week: midterm problems & answers, differentiation review
- Lecture next week
  - Will post video lectures for Wed & Fri
  - TA will go over material in detail in class and answer questions

Bias-Variance tradeoff – Intuition

- Model too simple: does not fit the data well
  - A *biased* solution
  - Simple = fewer features
  - Simple = more regularization

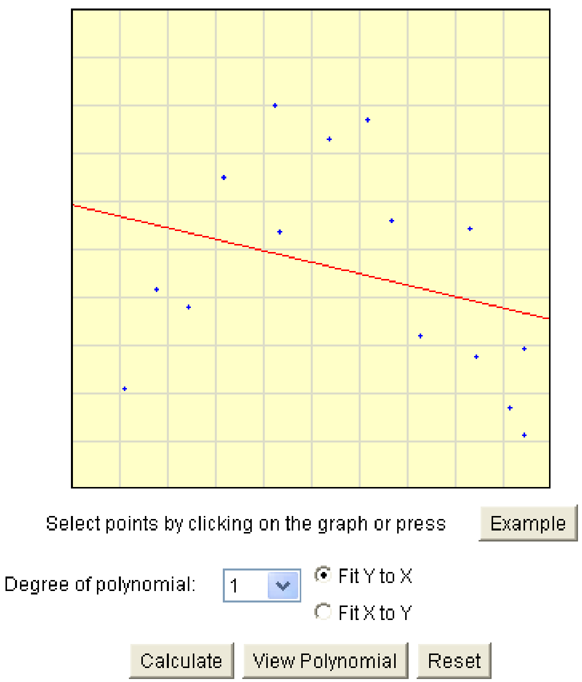


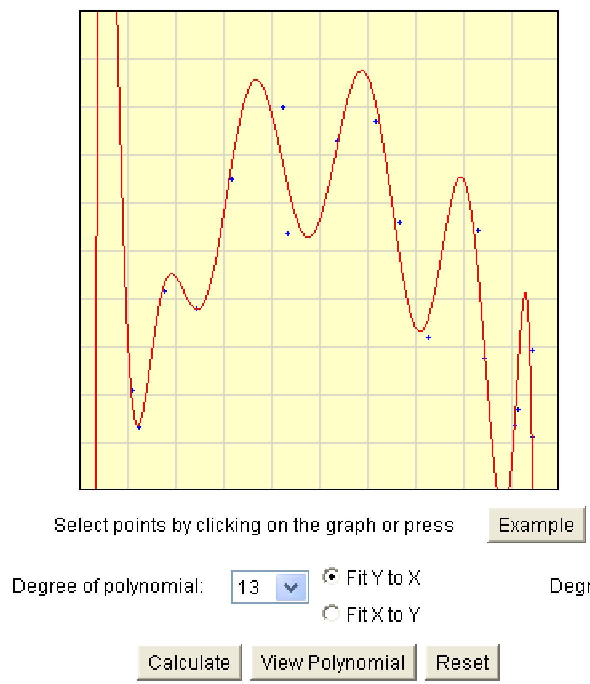
- Model too complex: small changes to the data, solution changes a lot
  - A *high-variance* solution
  - Complex = more features
  - Complex = less regularization

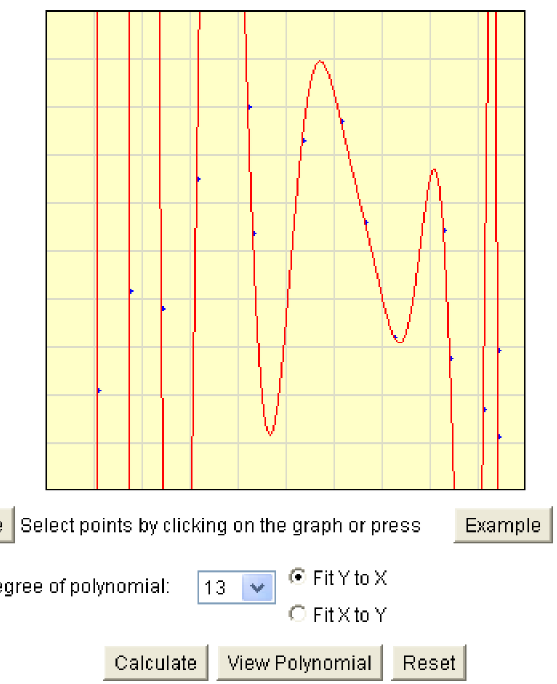


Bias-Variance Tradeoff

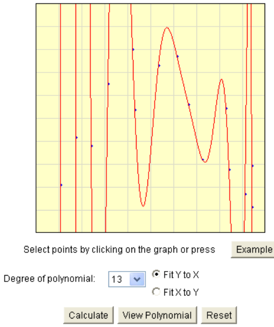
- Choice of hypothesis class introduces learning bias
  - More complex class → less bias
  - More complex class → more variance







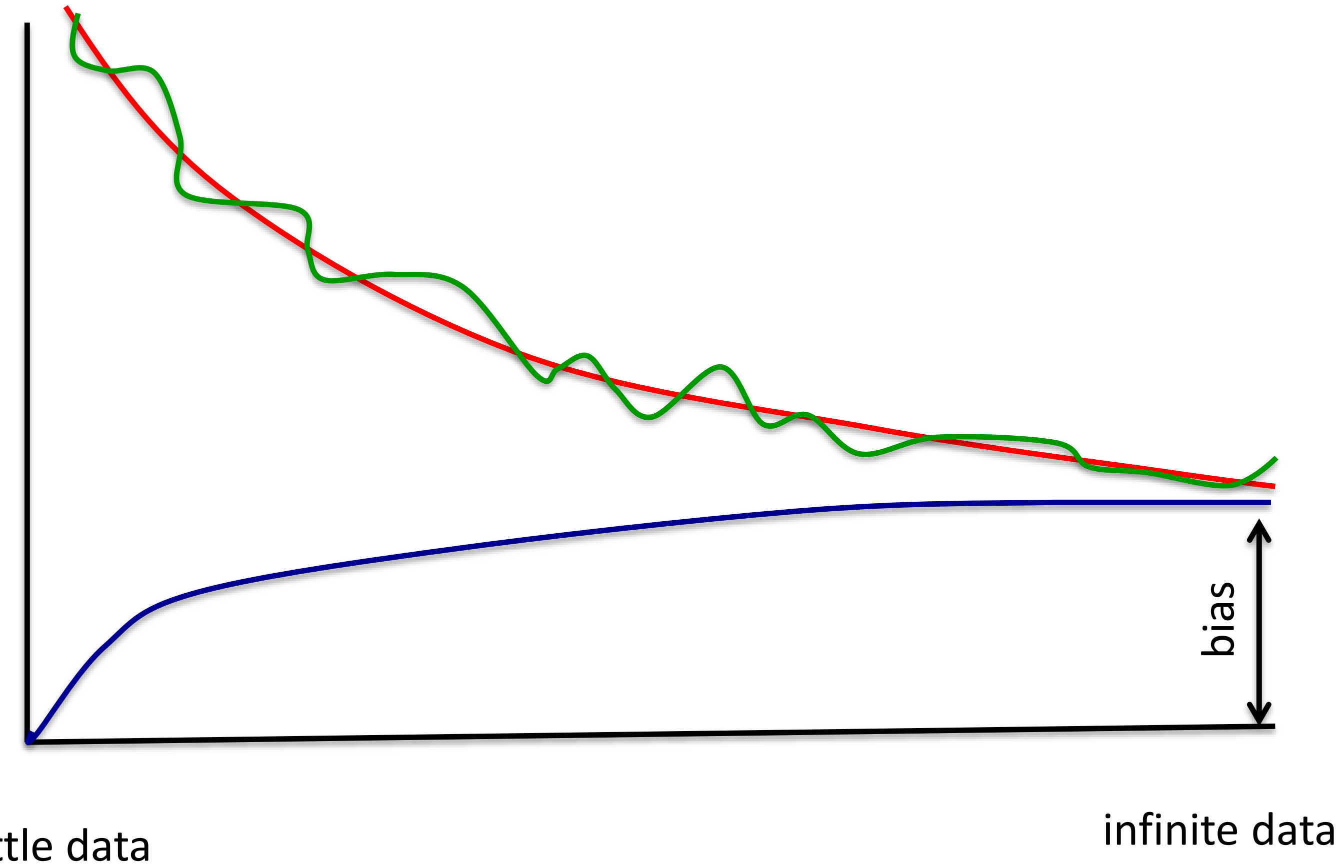
Error as a function of number of training examples for a fixed model complexity



$$\mathcal{E}_{\text{train}}(w) = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} (x_i \cdot w - y_i)^2$$

$$\mathcal{E}_{\text{true}}(w) = E_{p(x)} \left[(x_i \cdot w - y_i)^2 \right]$$

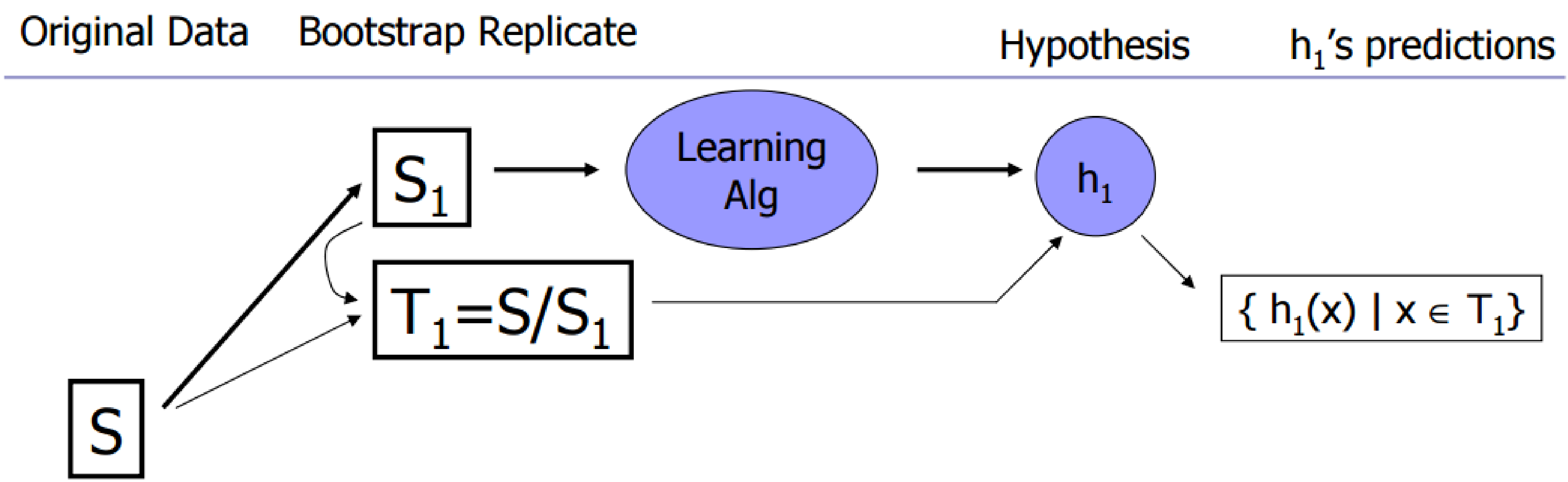
$$\mathcal{E}_{\text{test}}(w) = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (x_i \cdot w - y_i)^2$$



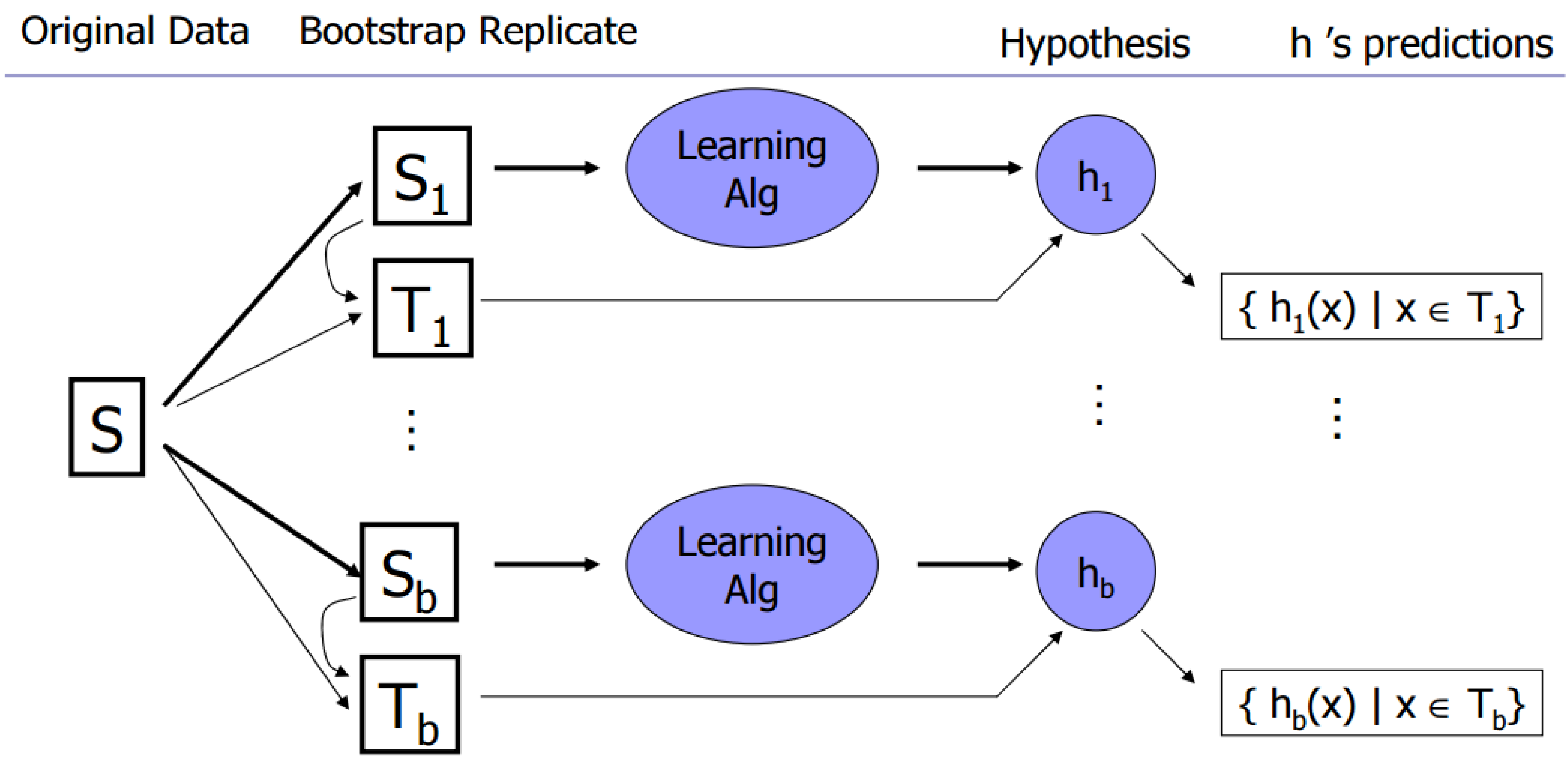
Measuring Bias and Variance

- In practice (unlike in theory), only *ONE training set D*
  - Simulate multiple training sets by **bootstrap replicates**
    - D' = {x | x is drawn at random with replacement from D }
    - |D'| = |D|

Estimating Bias / Variance



Estimating Bias / Variance



- Each S_i is bootstrap replicate
- $T_i = S / S_i$
- h_i = hypothesis, based on S_i

Average Response for each x_i

	x_1	…	x_r
$\in^? T_1$	$h_1(x_1)$	…	
$\in^? T_2$	--	…	$h_2(x_r)$
$\vdots$			
$\in^? T_b$	$h_b(x_1)$	…	$h_b(x_r)$

$\underline{h(x_1)} = 1/k_1 \, \Sigma \, h_i(x_1)$ … $\underline{h(x_r)} = 1/k_r \, \Sigma \, h_i(x_r)$

$$\underline{h(x_j)} = \Sigma_{\{i:\ x \in Ti\}} h_i(x_j) \,/\, ||\{i:\ x \in T_i\}||$$

Procedure for Measuring Bias and Variance

- Construct B bootstrap replicates of S
  $S_1, \ldots, S_B$
- Apply learning alg to each replicate S_b to obtain hypothesis h_b
- Let $T_b = S \setminus S_b$ = data points not in S_b (*out of bag* points)
- Compute predicted value
  $h_b(x)$
  for each $x \in T_b$

Estimating Bias and Variance

- For each $x \in S$,
  - observed response y
  - predictions $y_1, \ldots, y_k$
- Compute average prediction $\underline{h(x)} = ave_i \{y_i\}$
- Estimate bias: $\underline{h(x)} - y$
- Estimate variance:

$$\Sigma_{\{i:\, x \in Ti\}} \left(h_i(x) - \underline{h(x)} \right)^2 / (k-1)$$

- Assume noise is 0