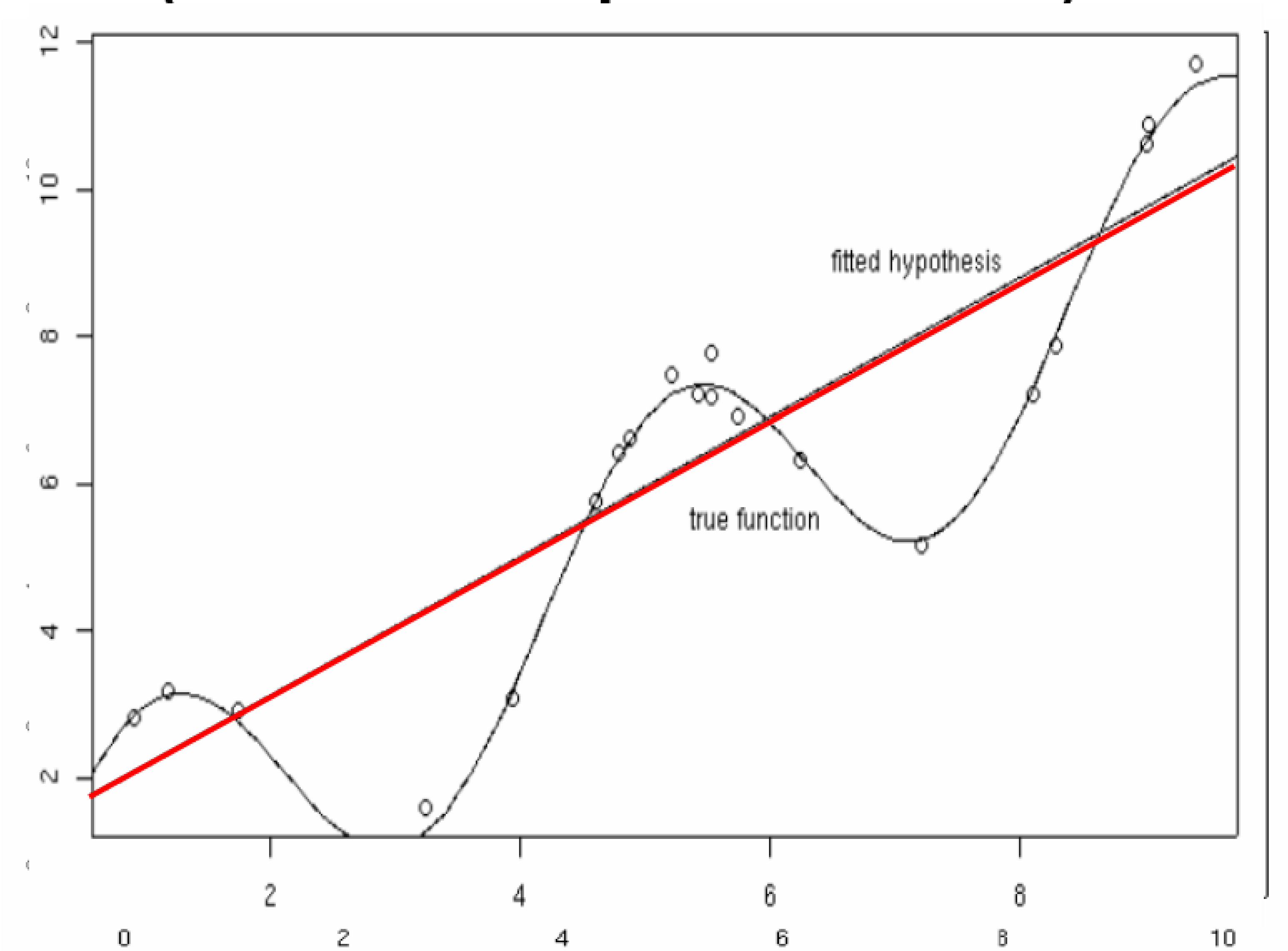
CSE 446
Learning Theory

Administrative

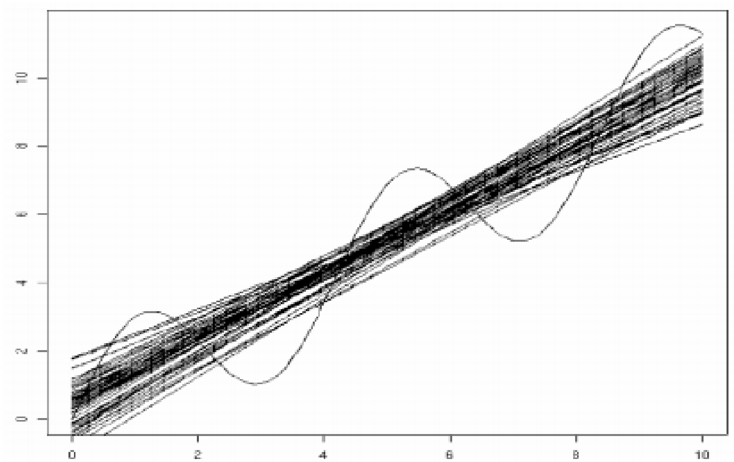
- Quiz section this week: midterm problems & answers, differentiation review
- Lecture this week
  - Video lectures posted for Wed & Fri
  - TA will go over material in detail in class and answer questions

50 fits (20 examples each)



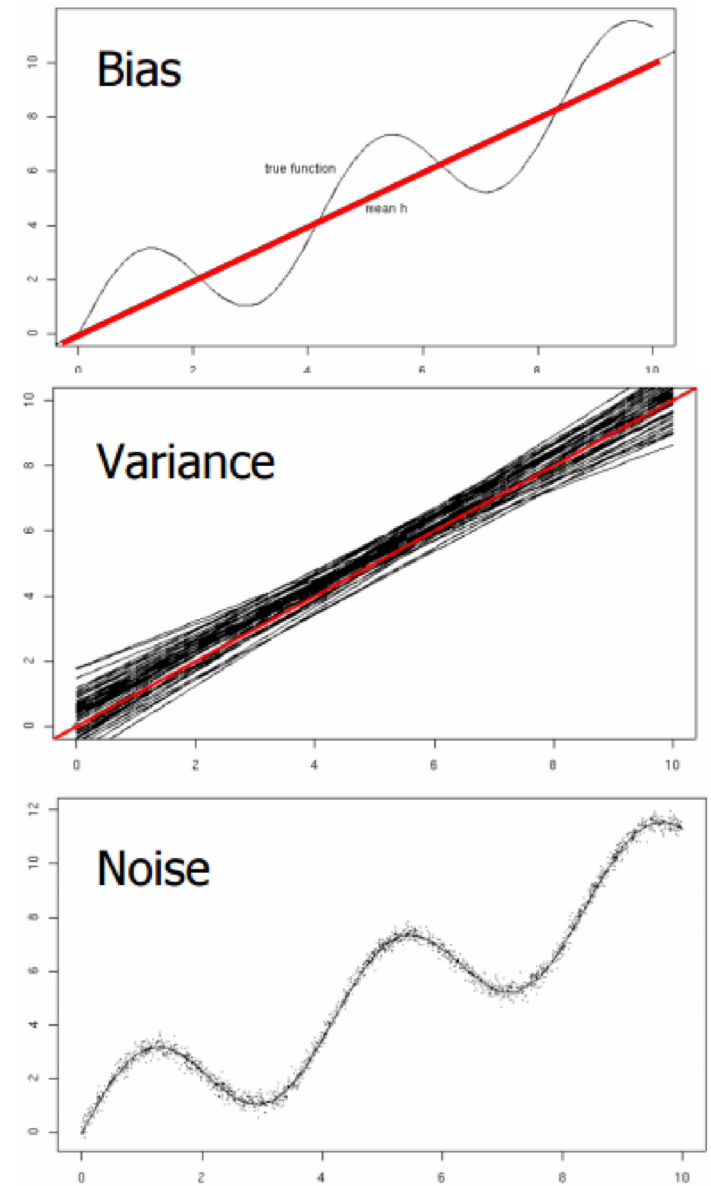
Slides from C Guestrin, T Dietterich, R Parr, N Ray, R Greiner

Bias, Variance, Noise

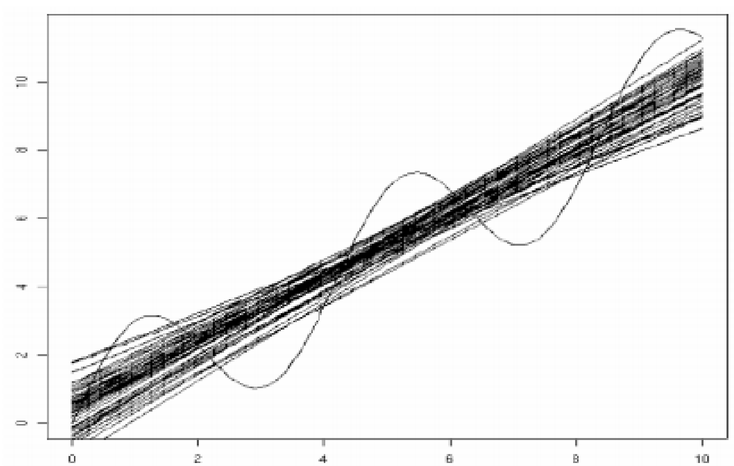


50 fits (20 examples each)

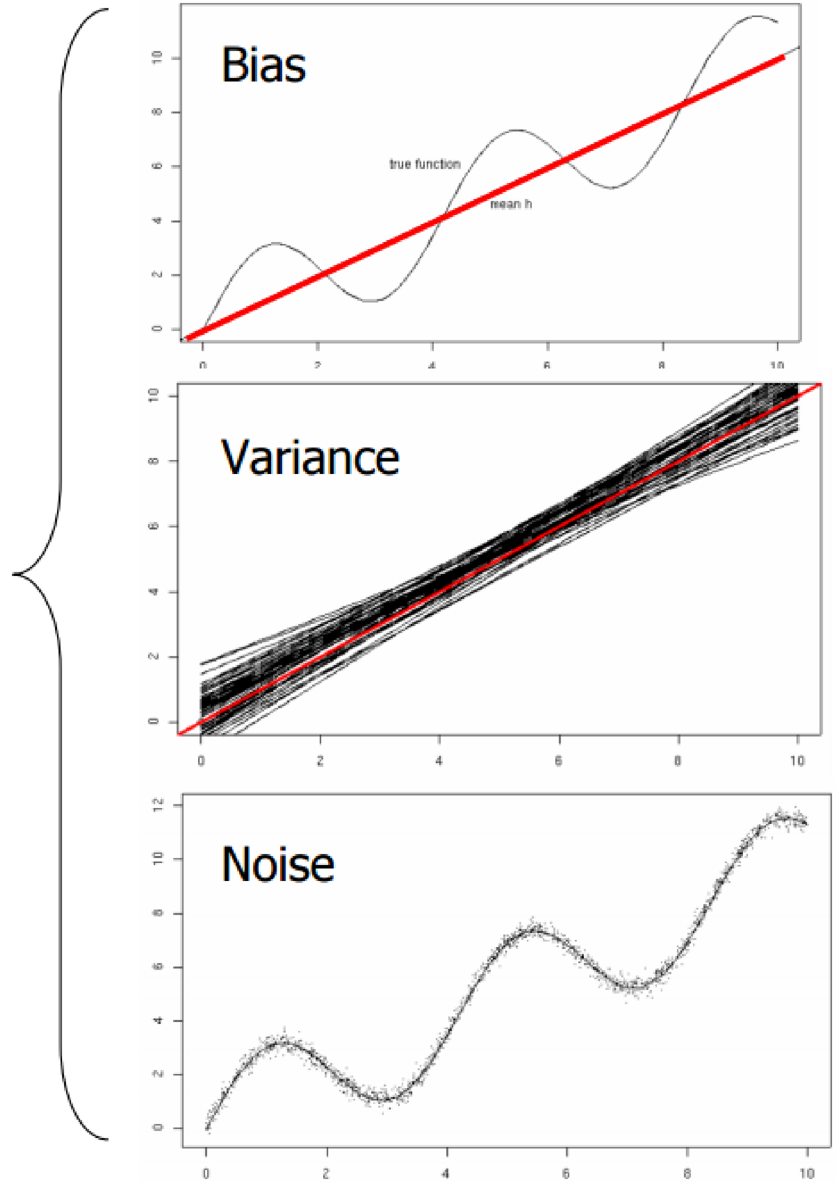
=



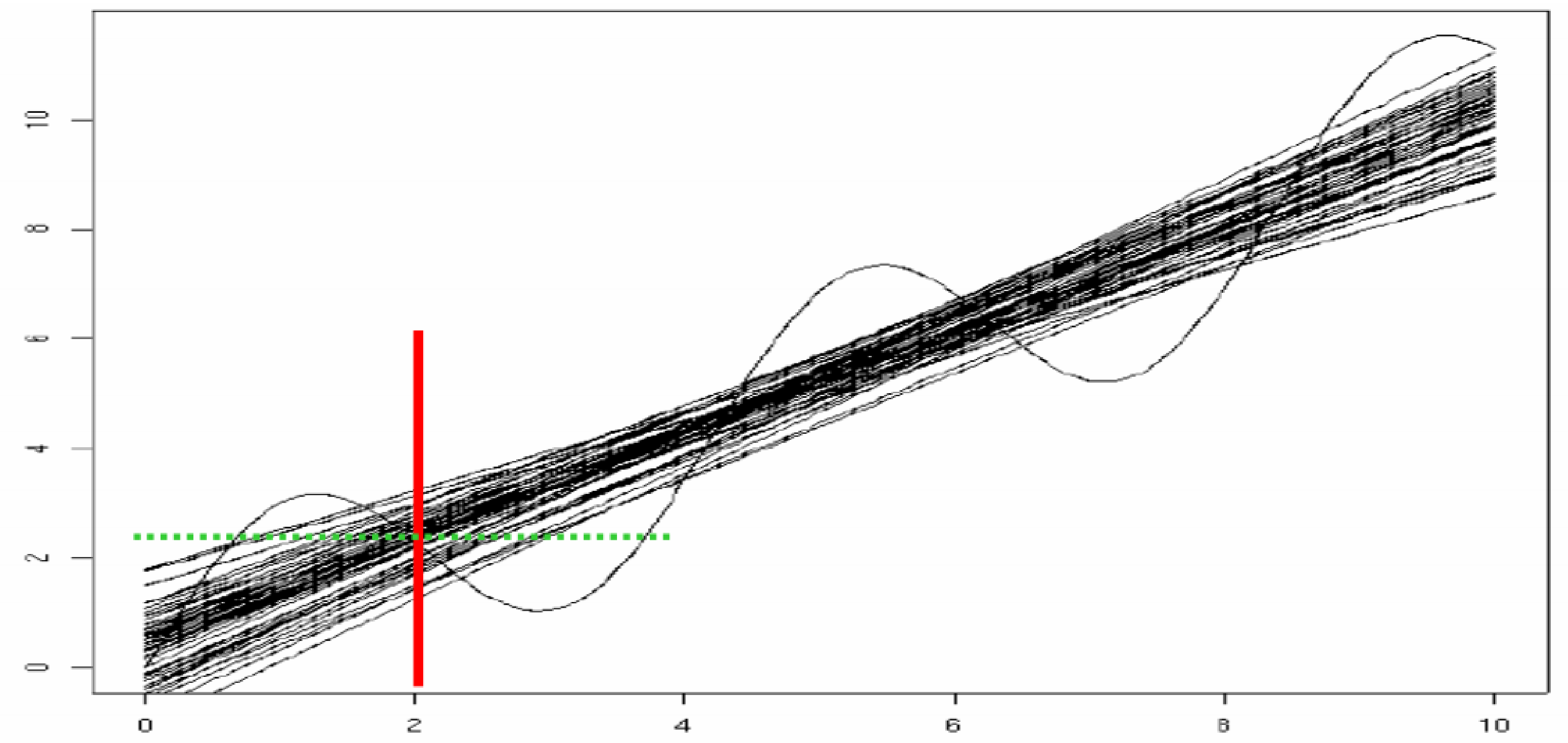
Bias, Variance, Noise



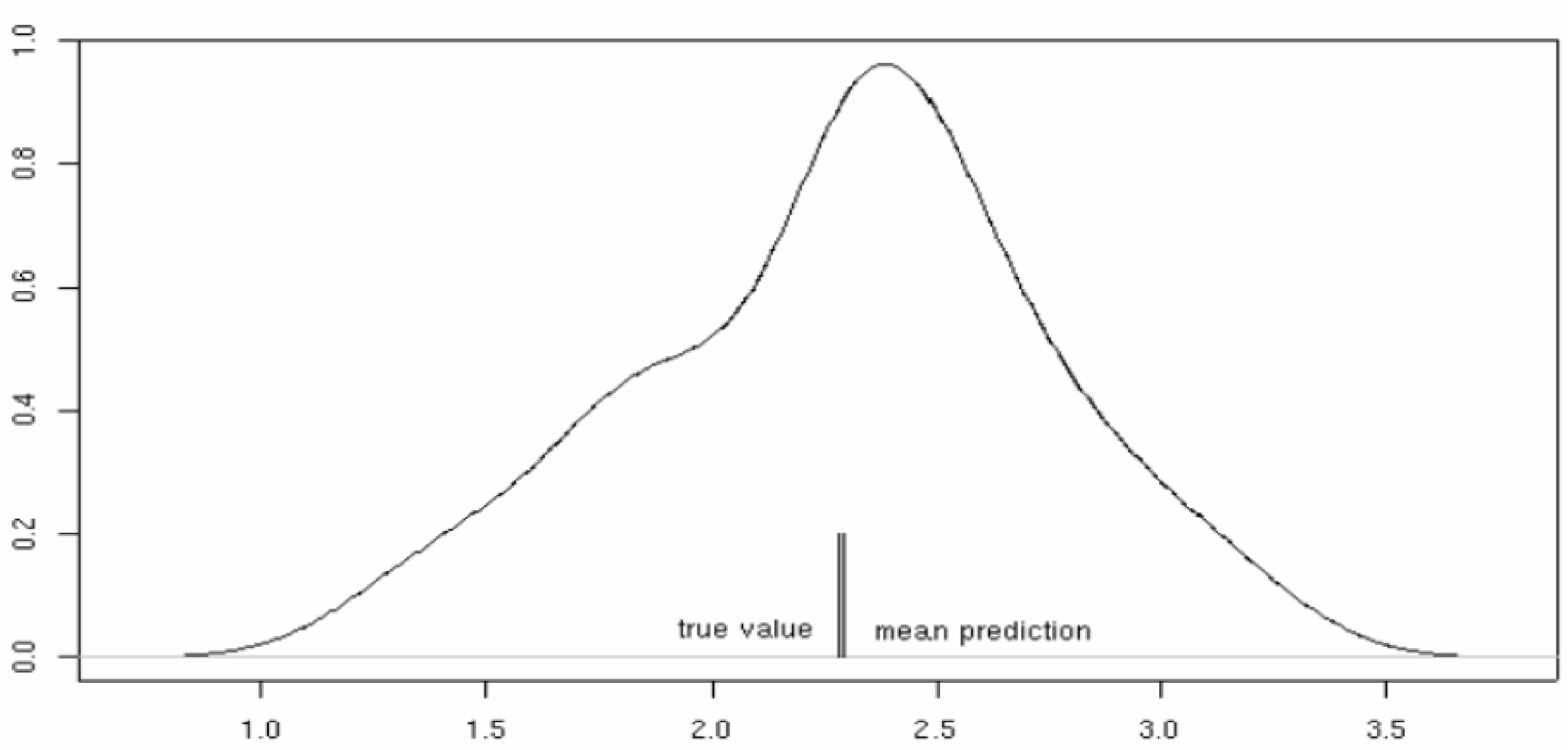
50 fits (20 examples each)



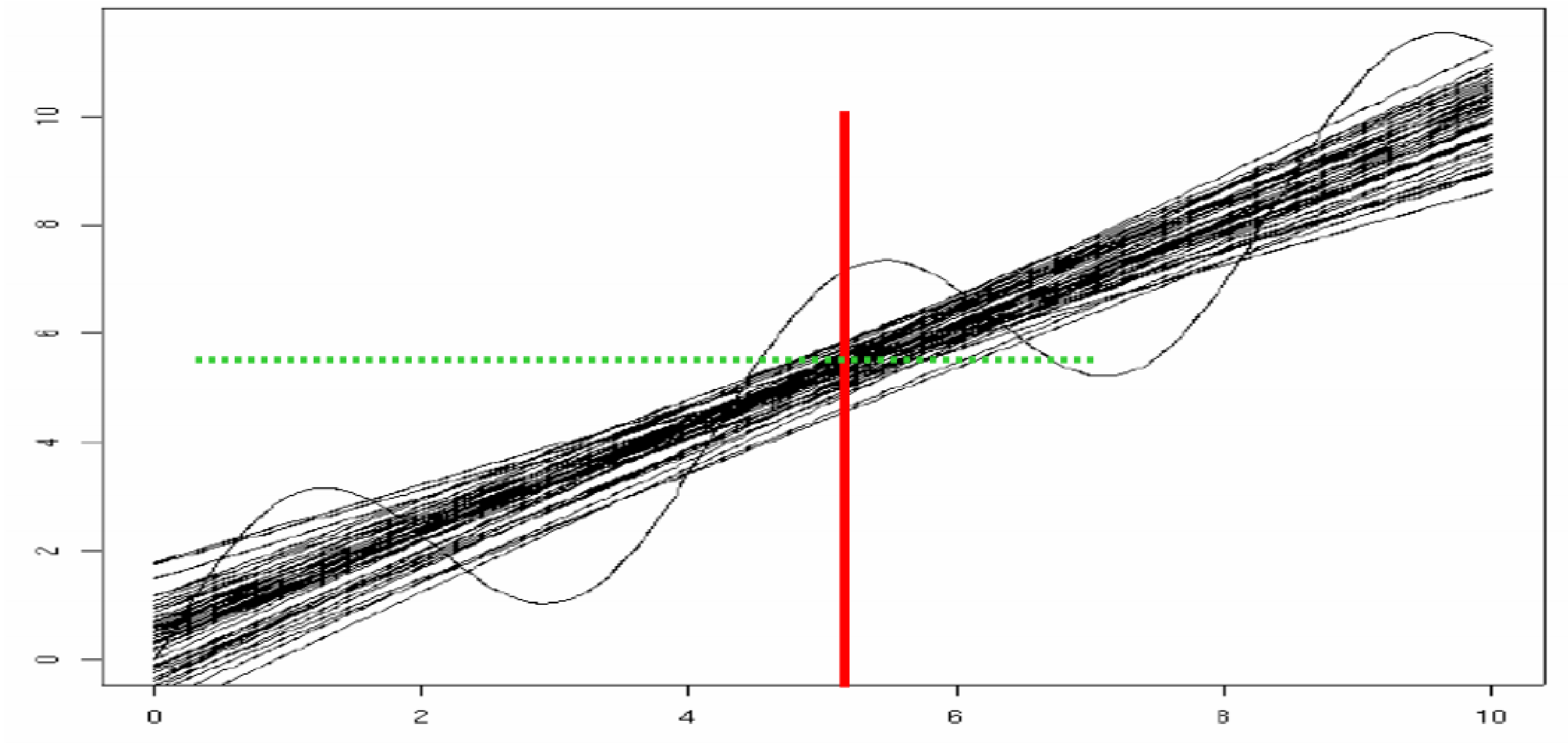
50 fits (20 examples each)



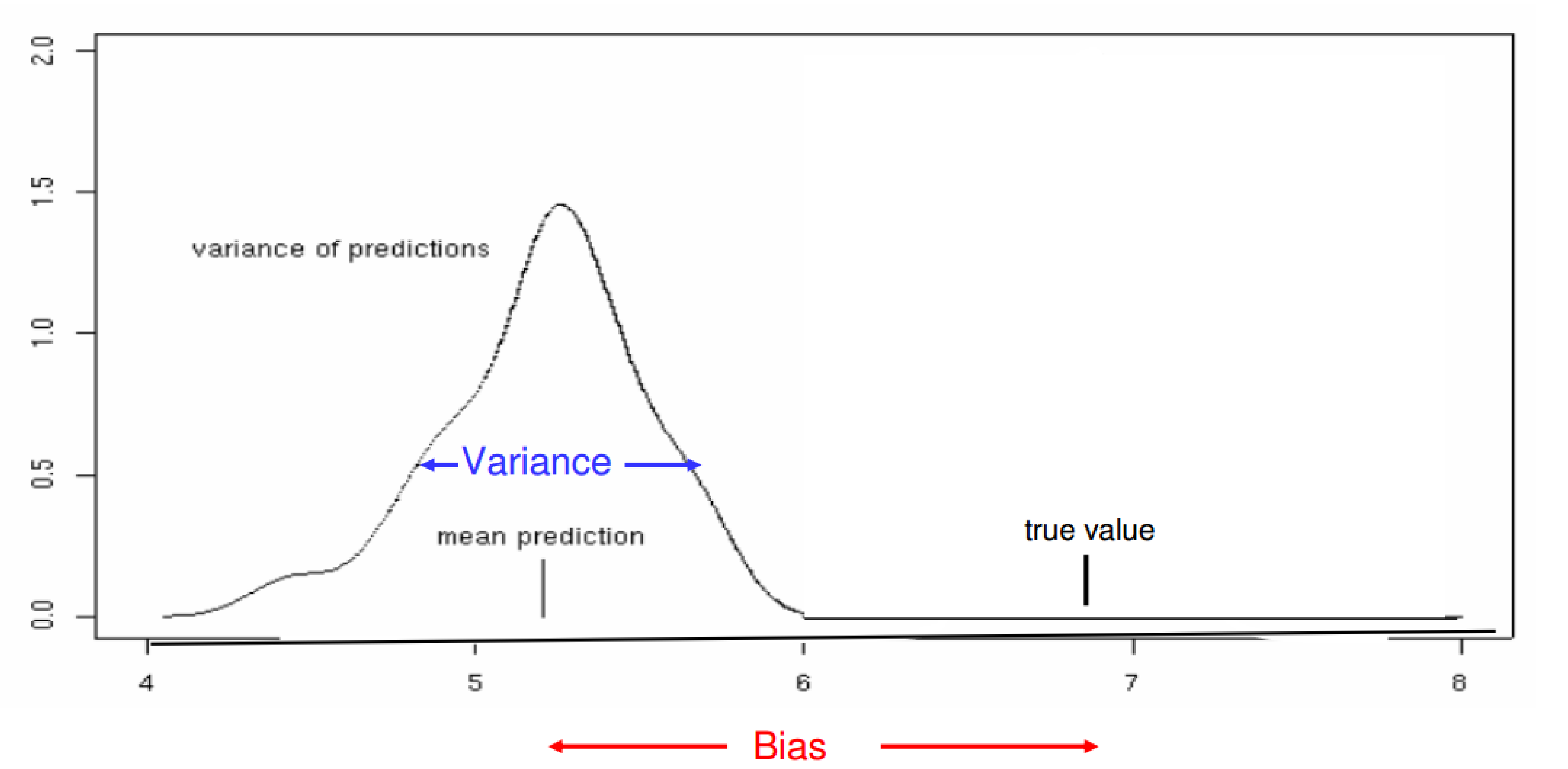
Predictions at x=2.0



50 fits (20 examples each)



Predictions at x=5.0

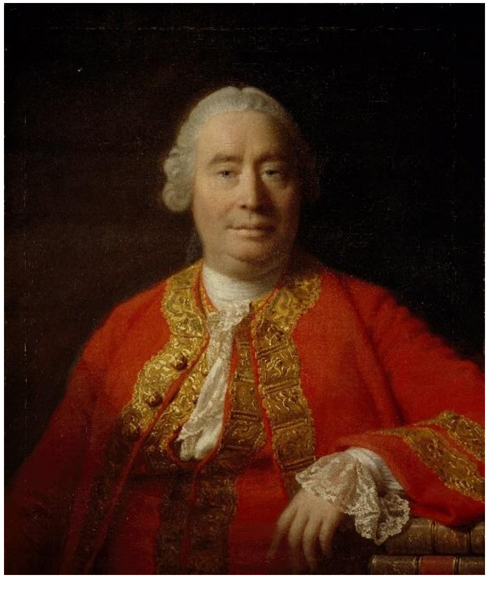


Inductive Bias

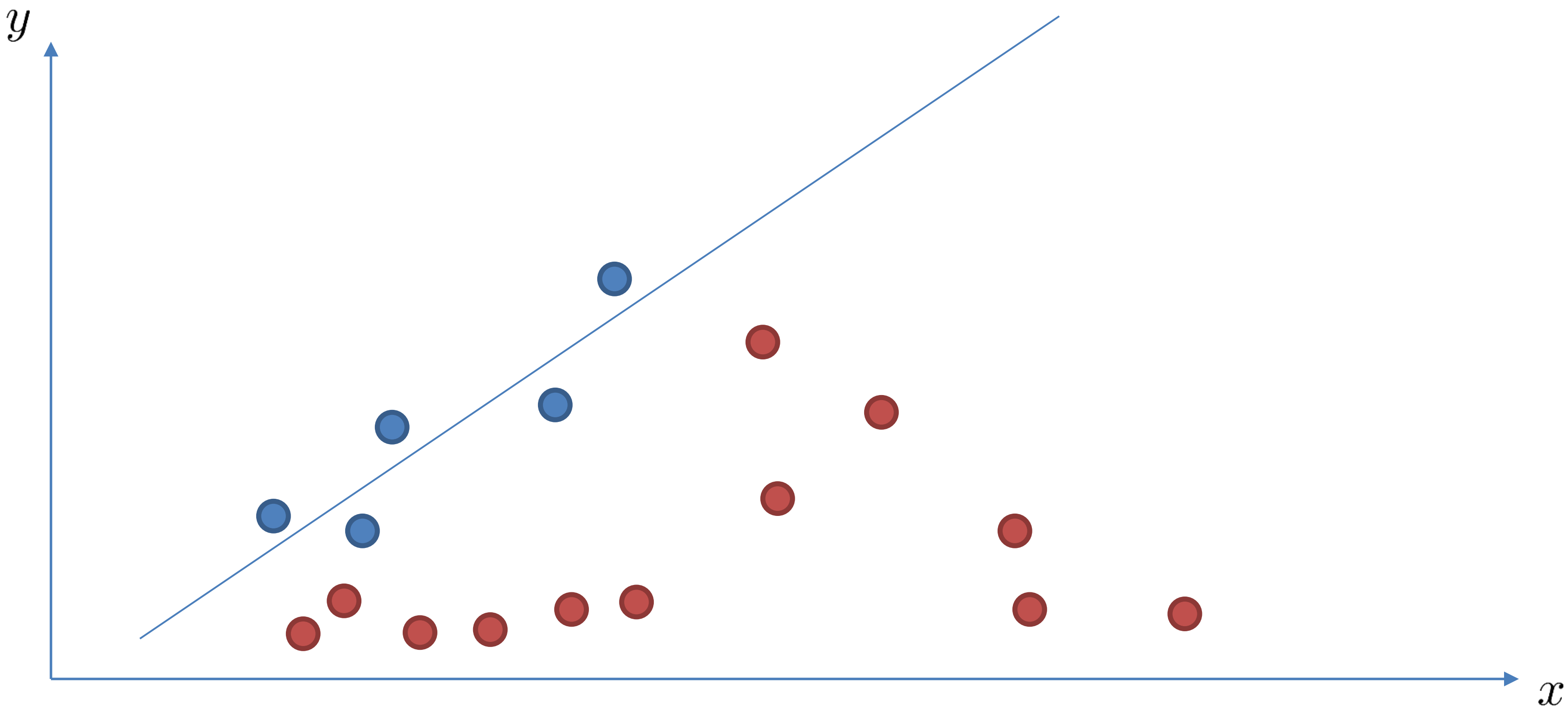
Inductive Bias

Even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience.

- David Hume, 1739



Inductive Bias



No Free Lunch Theorem

$\mathcal{D}$ – training set

$g(\mathbf{x})$ – the true function that gives rise to the data

$\hat{f}(\mathbf{x})$ – hypothesis predicted by our learning algorithm

$A(\mathcal{D})$ – learning algorithm (takes in $\mathcal{D}$ and produces $\hat{f}(\mathbf{x})$)

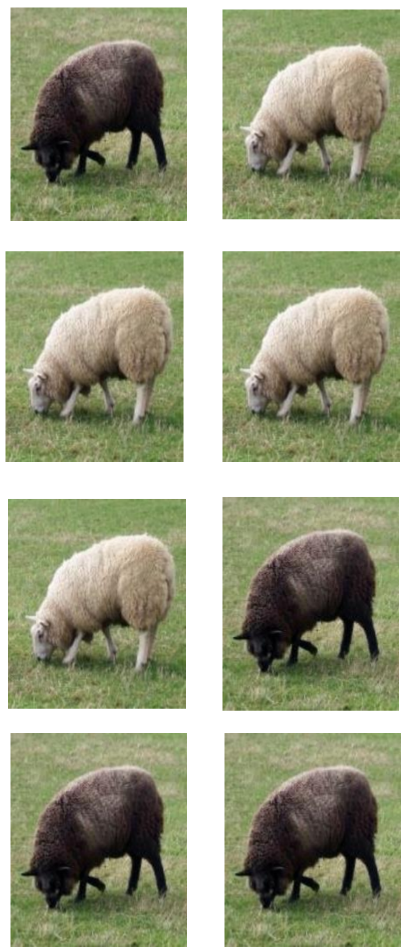
$\mathcal{E}_{\text{gen},\mathcal{D}}(\hat{f})$ – error on all data not in the set $\mathcal{D}$, depends on hypothesis $\hat{f}$

$$E_{p(g)}[\mathcal{E}_{\text{gen},\mathcal{D}}(A(\mathcal{D}))] = E_{p(g)}[\mathcal{E}_{\text{gen},\mathcal{D}}(A'(\mathcal{D}))].$$

If all true functions are equally likely, no learning algorithm is better than any other.

Example

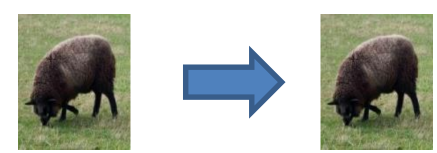
possible worlds
(consisting of 2 sheep)

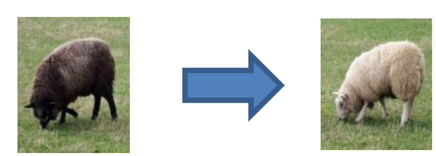


possible algorithms

guess black  1/2

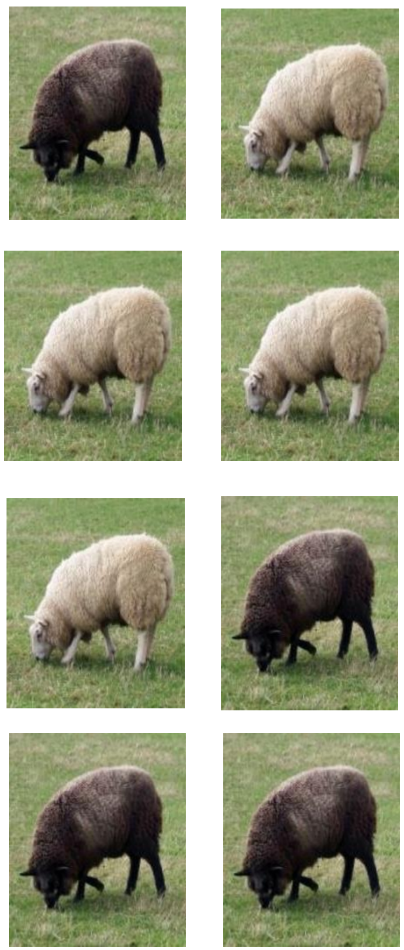
guess white  1/2

guess same  1/2

guess different  1/2

Add Knowledge: Single-Color Sheep

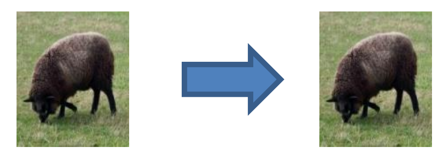
possible worlds
(consisting of 2 sheep)

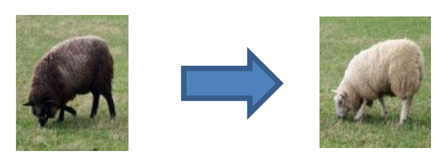


possible algorithms

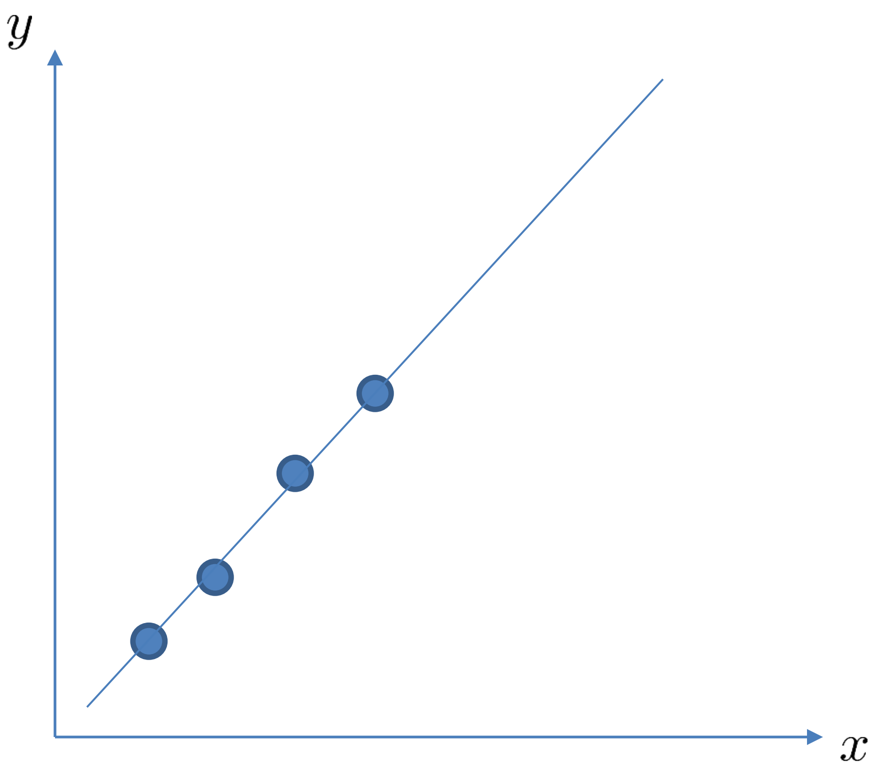
guess black  1/2

guess white  1/2

guess same  1!

guess different  0

Representation Matters



$x = 2 \;\; y = 2$

$x = 3 \;\; y = 3$

$x = 6 \;\; y = 6$

$x = 12 \;\; y = ?$

$x = (0, 1, 0, 0) \;\; y = 2$

$x = (1, 1, 0, 0) \;\; y = 3$

$x = (0, 1, 1, 0) \;\; y = 6$

$x = (0, 1, 0, 1) \;\; y = ?$

Takeaways

- Cross-validation
  - Good way to choose representation, hyperparameters, etc.
  - Still a learning algorithm, not immune to NFL
- Representation vs algorithm
  - NFL doesn’t distinguish these
  - Some algorithms really do work very very well in many many cases… with the right representation & regularization
- Domain knowledge
  - Hugely important, helps a lot
  - But be careful not to introduce too much bias…