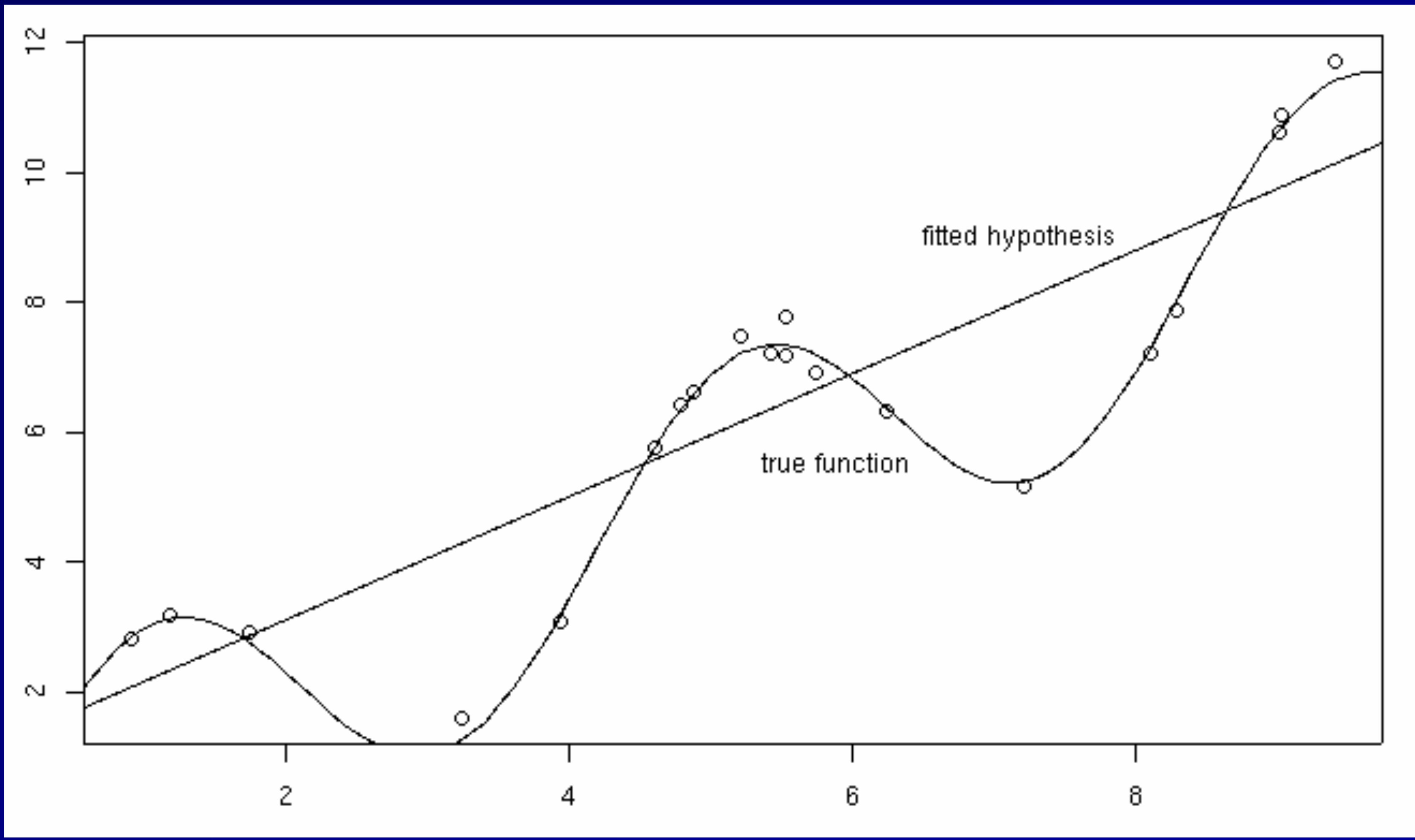
Bias-Variance Analysis in Regression

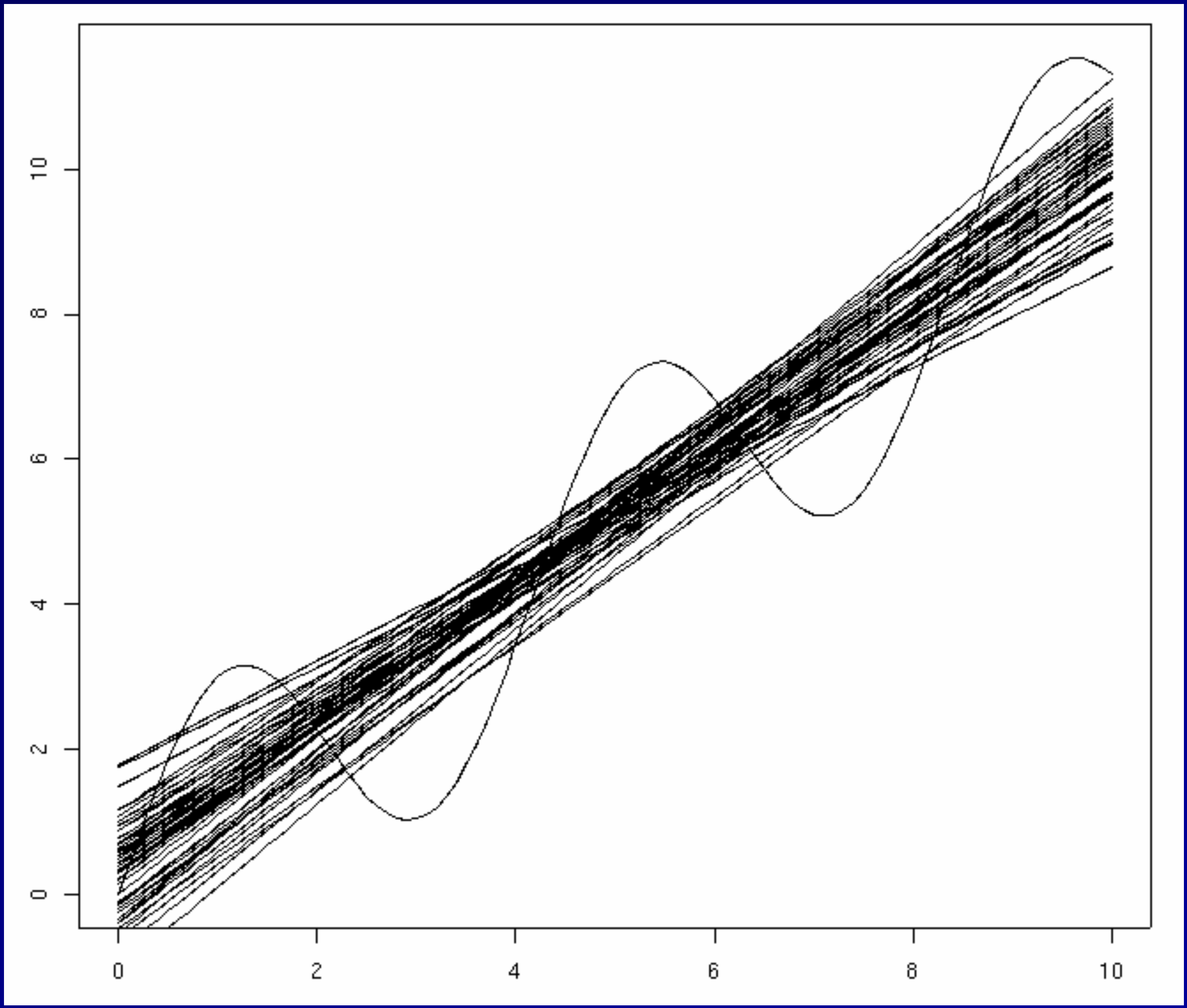
- True function is $y = f(x) + \varepsilon$
  - where ε is normally distributed with zero mean and standard deviation σ.
- Given a set of training examples, $\{(x_i, y_i)\}$, we fit an hypothesis $h(x) = w \cdot x + b$ to the data to minimize the squared error

$$\Sigma_i [y_i - h(x_i)]^2$$

Example: 20 points
$y = x + 2\sin(1.5x) + N(0,0.2)$



50 fits (20 examples each)



Bias-Variance Analysis

- Now, given a new data point x* (with observed value $y^* = f(x^*) + \varepsilon$), we would like to understand the expected prediction error

$$E[\,(y^* - h(x^*))^2\,]$$

Classical Statistical Analysis

- Imagine that our particular training sample S is drawn from some population of possible training samples according to P(S).
- Compute $E_P [(y^* - h(x^*))^2]$
- Decompose this into “bias”, “variance”, and “noise”

Lemma

- Let Z be a random variable with probability distribution P(Z)
- Let $\underline{\underline{Z}} = E_P[\, Z \,]$ be the average value of Z.
- Lemma: $E[\, (Z - \underline{\underline{Z}})^2 \,] = E[Z^2] - \underline{\underline{Z}}^2$

$$E[\, (Z - \underline{\underline{Z}})^2 \,] = E[\, Z^2 - 2\, Z\, \underline{\underline{Z}} + \underline{\underline{Z}}^2 \,]$$
$$= E[Z^2] - 2\, E[Z]\, \underline{\underline{Z}} + \underline{\underline{Z}}^2$$
$$= E[Z^2] - 2\, \underline{\underline{Z}}^2 + \underline{\underline{Z}}^2$$
$$= E[Z^2] - \underline{\underline{Z}}^2$$

- Corollary: $E[Z^2] = E[\, (Z - \underline{\underline{Z}})^2 \,] + \underline{\underline{Z}}^2$

Bias-Variance-Noise Decomposition

$E[\,(h(x^*) - y^*)^2\,] = E[\,h(x^*)^2 - 2\,h(x^*)\,y^* + y^{*2}\,]$

$= E[\,h(x^*)^2\,] - 2\,E[\,h(x^*)\,]\,E[y^*] + E[y^{*2}]$

$= E[\,(h(x^*) - \underline{h(x^*)})^2\,] + \underline{h(x^*)}^2$ (lemma)

$- 2\,\underline{h(x^*)}\,f(x^*)$

$+ E[\,(y^* - f(x^*))^2\,] + f(x^*)^2$ (lemma)

$= E[\,(h(x^*) - \underline{h(x^*)})^2\,] +$ [variance]

$(\underline{h(x^*)} - f(x^*))^2 +$ [bias^2]

$E[\,(y^* - f(x^*))^2\,]$ [noise]

Derivation (continued)

$E[\,(h(x^*) - y^*)^2\,] =$

$\quad = E[\,(h(x^*) - \underline{h(x^*)})^2\,] +$

$\quad\quad (\underline{h(x^*)} - f(x^*))^2 +$

$\quad\quad E[\,(y^* - f(x^*))^2\,]$

$\quad = Var(h(x^*)) + Bias(h(x^*))^2 + E[\,\varepsilon^2\,]$

$\quad = Var(h(x^*)) + Bias(h(x^*))^2 + \sigma^2$

Expected prediction error = Variance + Bias^2 + Noise^2

Bias, Variance, and Noise

- Variance: $E[\, (h(x^*) - \underline{h(x^*)})^2 \,]$

  Describes how much $h(x^*)$ varies from one training set S to another

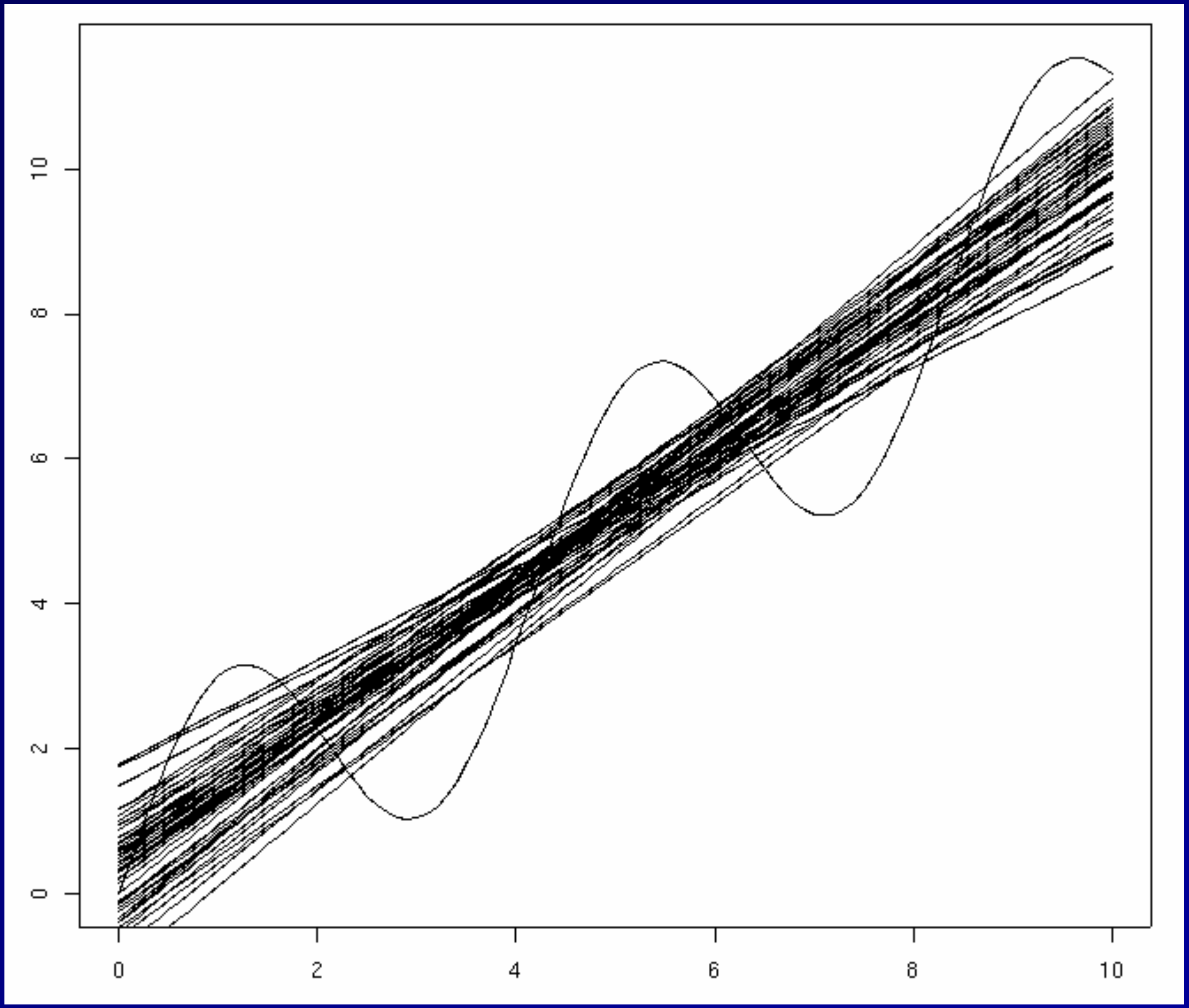
- Bias: $[\underline{h(x^*)} - f(x^*)]$

  Describes the <u>average</u> error of $h(x^*)$.

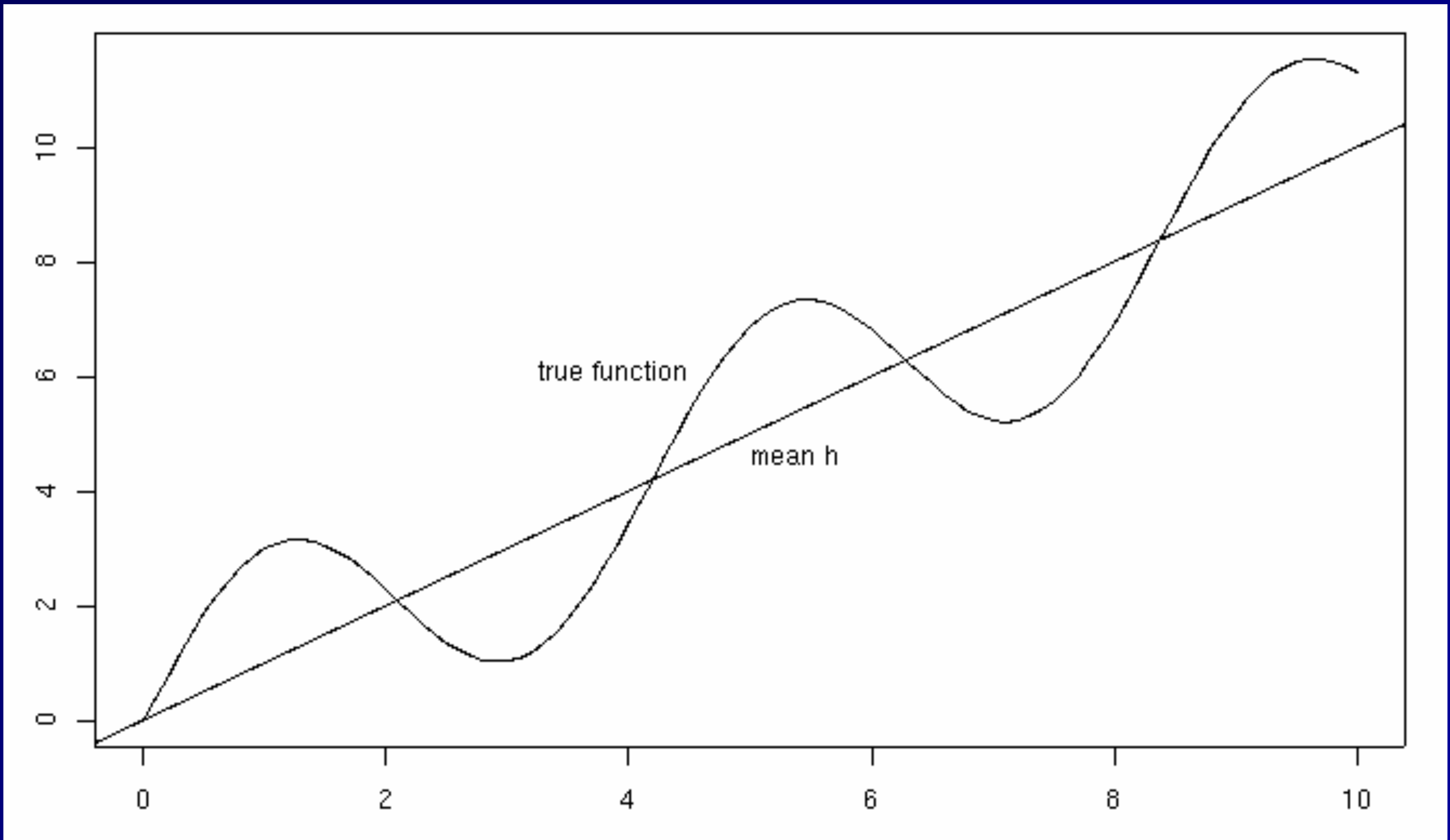
- Noise: $E[\, (y^* - f(x^*))^2 \,] = E[\varepsilon^2] = \sigma^2$

  Describes how much y^* varies from $f(x^*)$

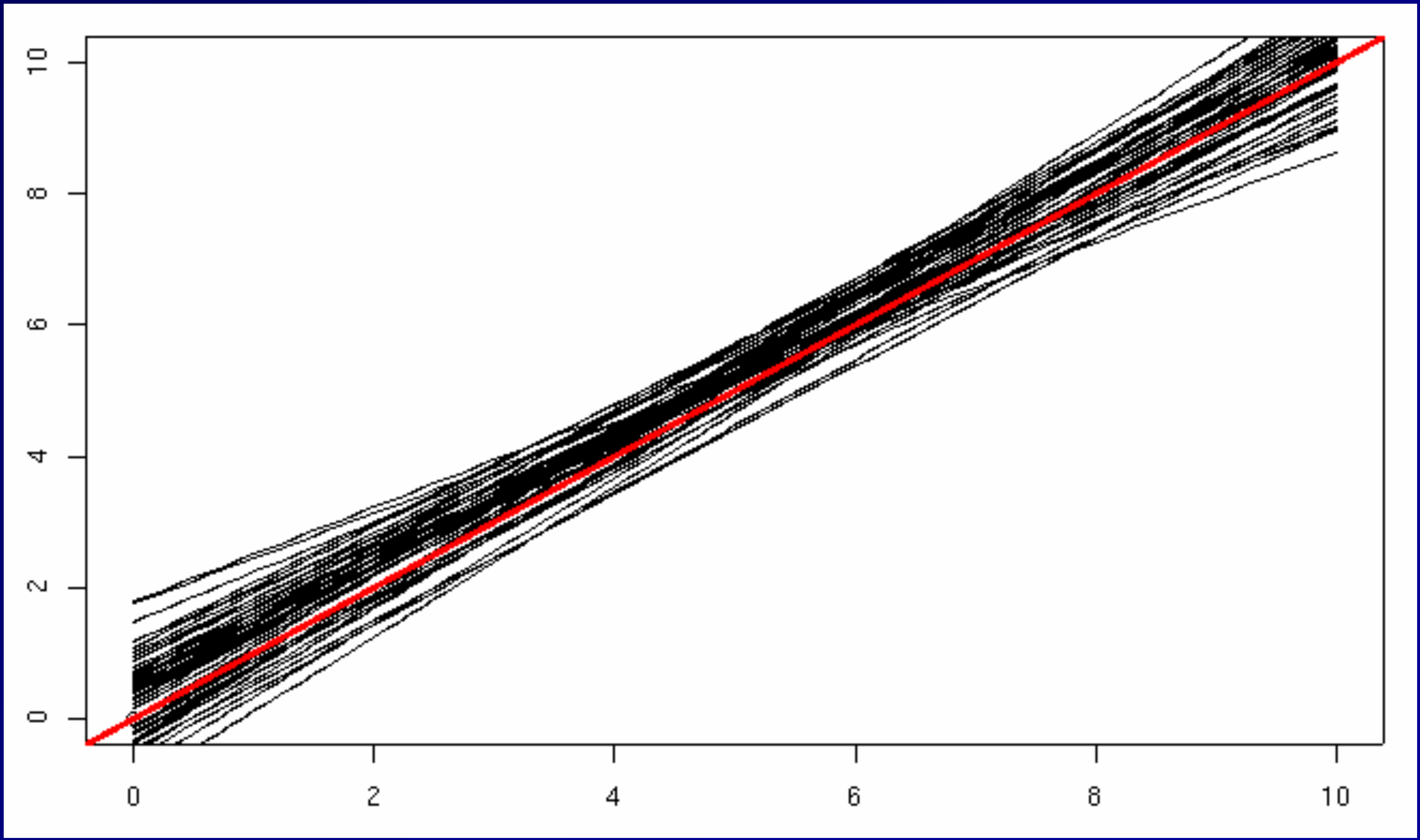
50 fits (20 examples each)



Bias



Variance



Noise