10-315
Introduction to ML

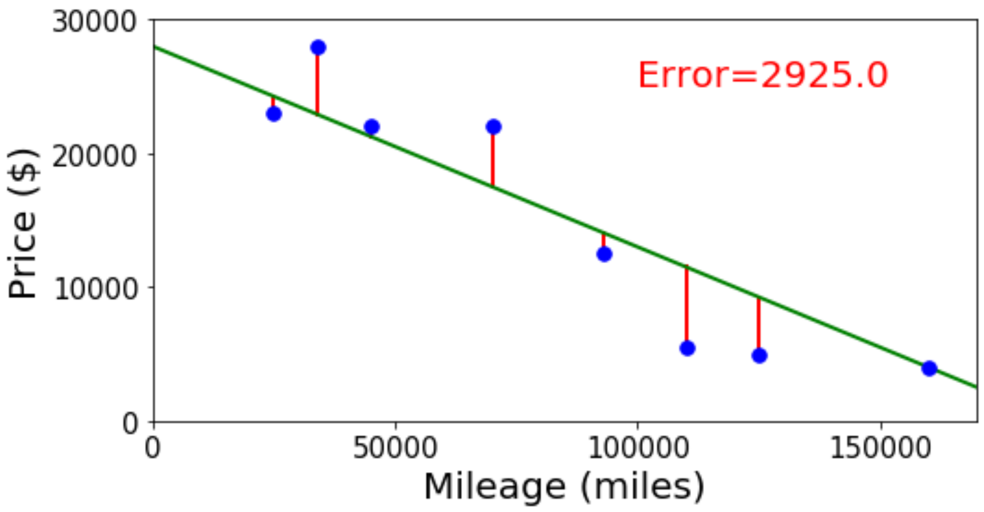
Feature Engineering

Instructor: Pat Virtue

Previous Poll

Linear regression. What is linear about it?

Select all that apply.



A. Always fits a linear (or affine) shape to the data

B. Linear objective function with respect to the input

C. Linear objective function with respect to the parameters

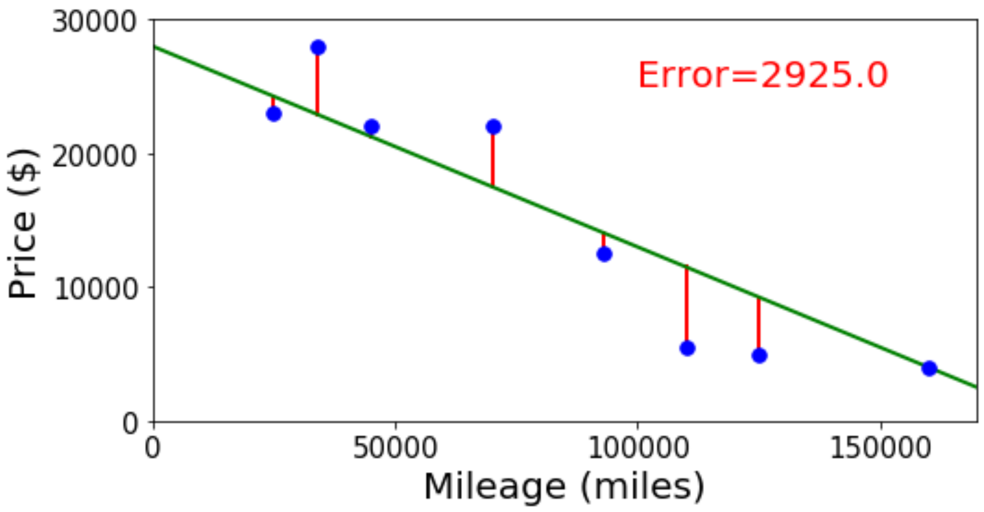
D. None of the above

$$y = mx + b$$
$$y = wx + b$$
$$y = w_1 x + w_0$$
$$y = \theta_1 x + \theta_0$$

Previous Poll

Linear regression. What is linear about it?

Select all that apply.



A. Always fits a linear (or affine) shape to the data

B. Linear objective function with respect to the input

C. Linear objective function with respect to the parameters

D. None of the above

$\vec{x}^T\theta$

$$\theta = \begin{bmatrix} b \\ w_1 \\ w_2 \end{bmatrix} \quad \vec{x} = \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix}$$

$y = mx + b$

$y = wx + b$

$y = w_1 x + w_0$

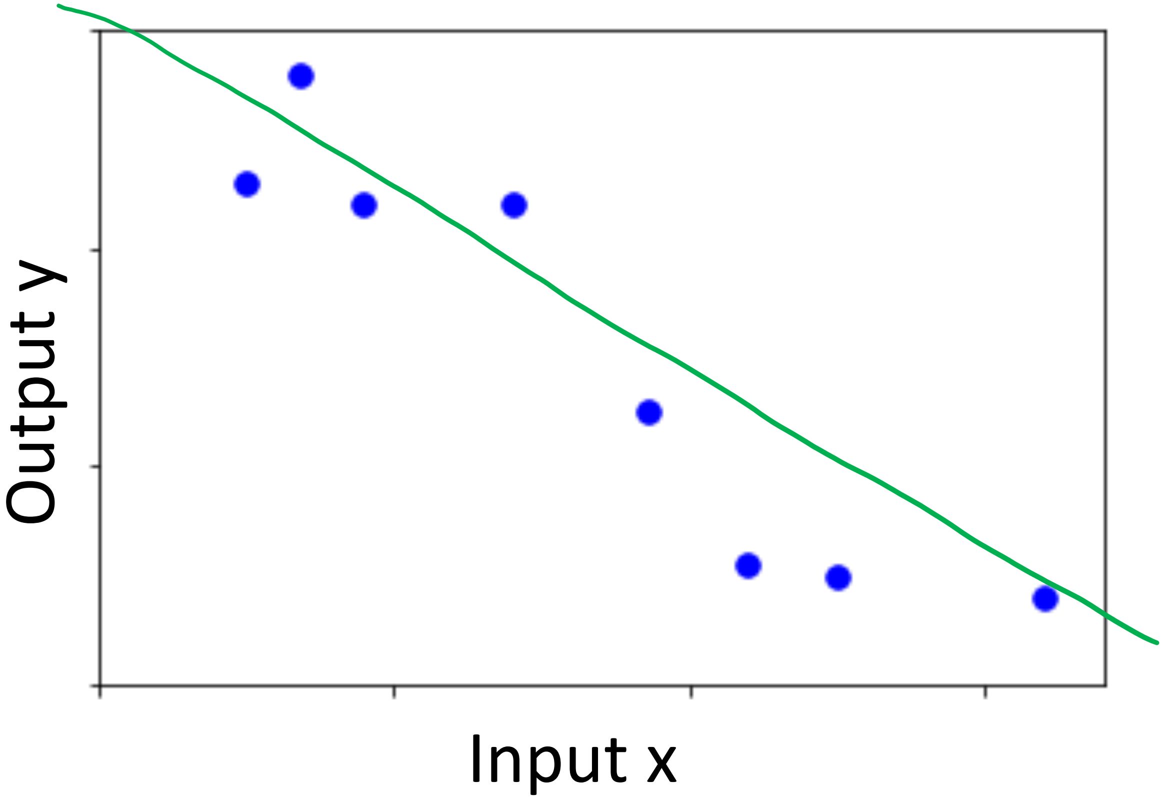
$y = \theta_1 x + \theta_0$

$y = w_1 x_1 + w_2 x_2 + b$

$y = w_2 \underline{x_1^2} + w_1 x_1 + b$

Linear Data

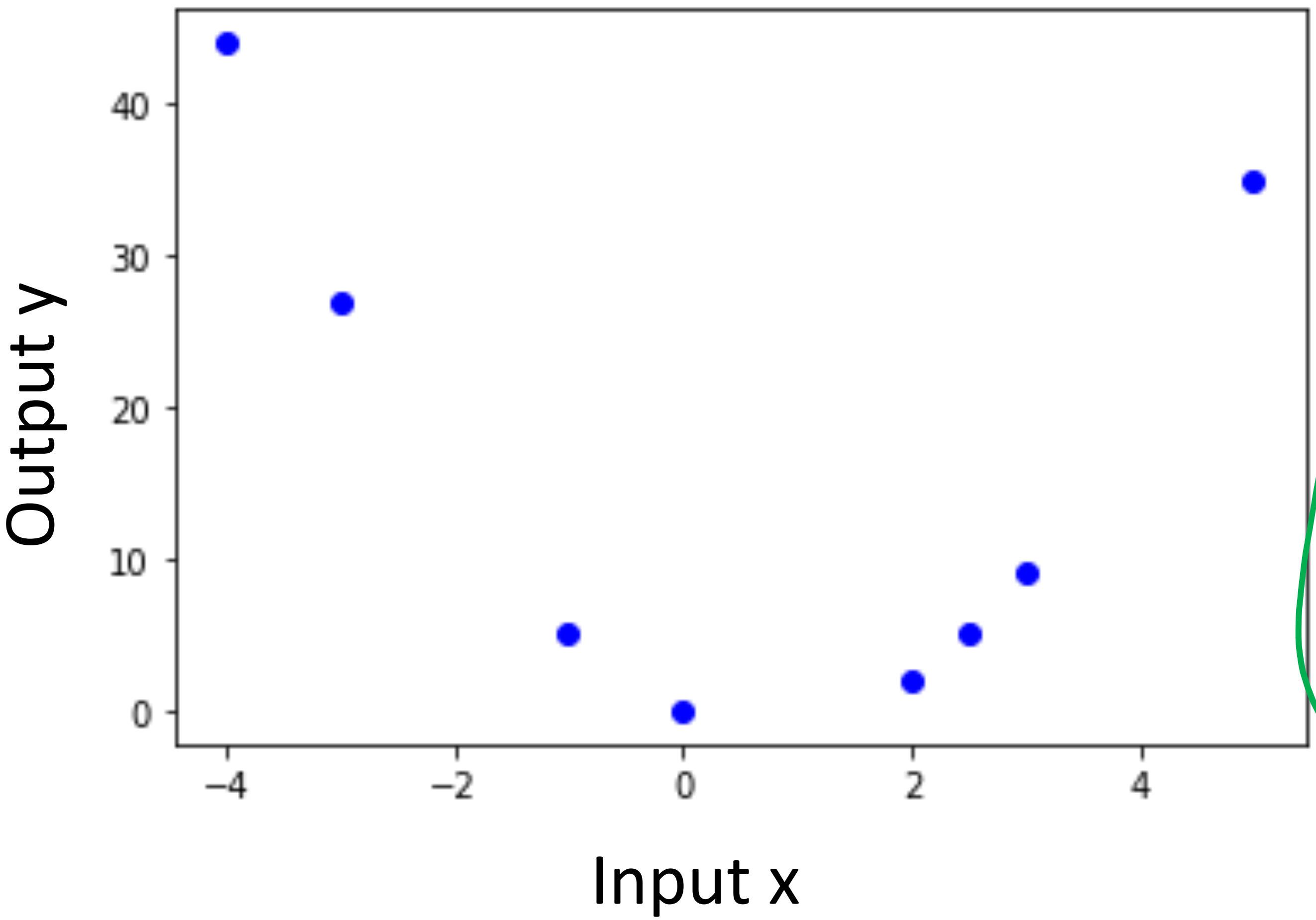
Regression on simple linear dataset



Non-linear Data

$D_1 \rightarrow D_2$

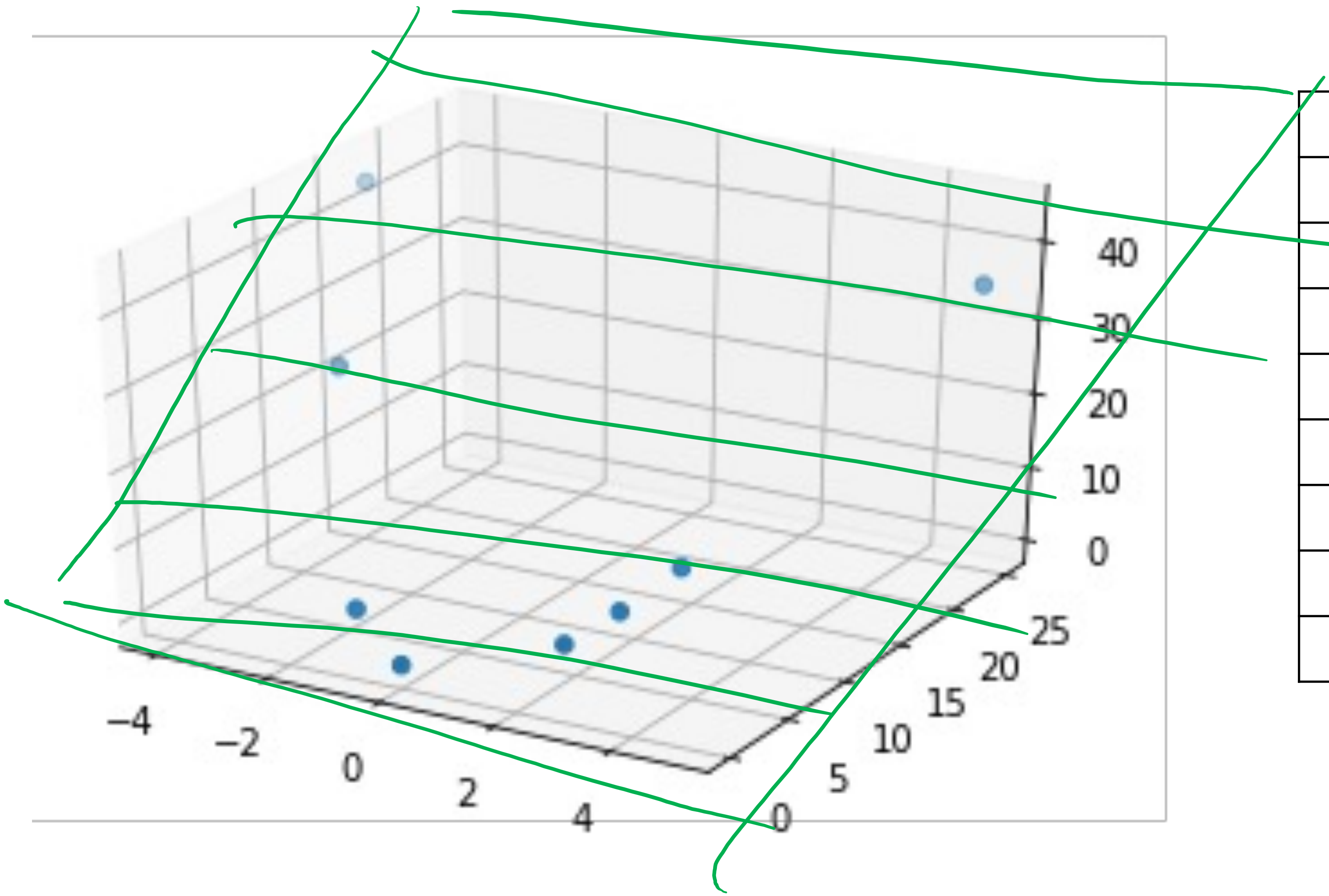
Linear regression on polynomial data?



y	x_1
44	-4
27	-3
5	-1
0	0
2	2
5	2.5
9	3
35	5

Non-linear Data

Regression on simple linear dataset



x_1^2

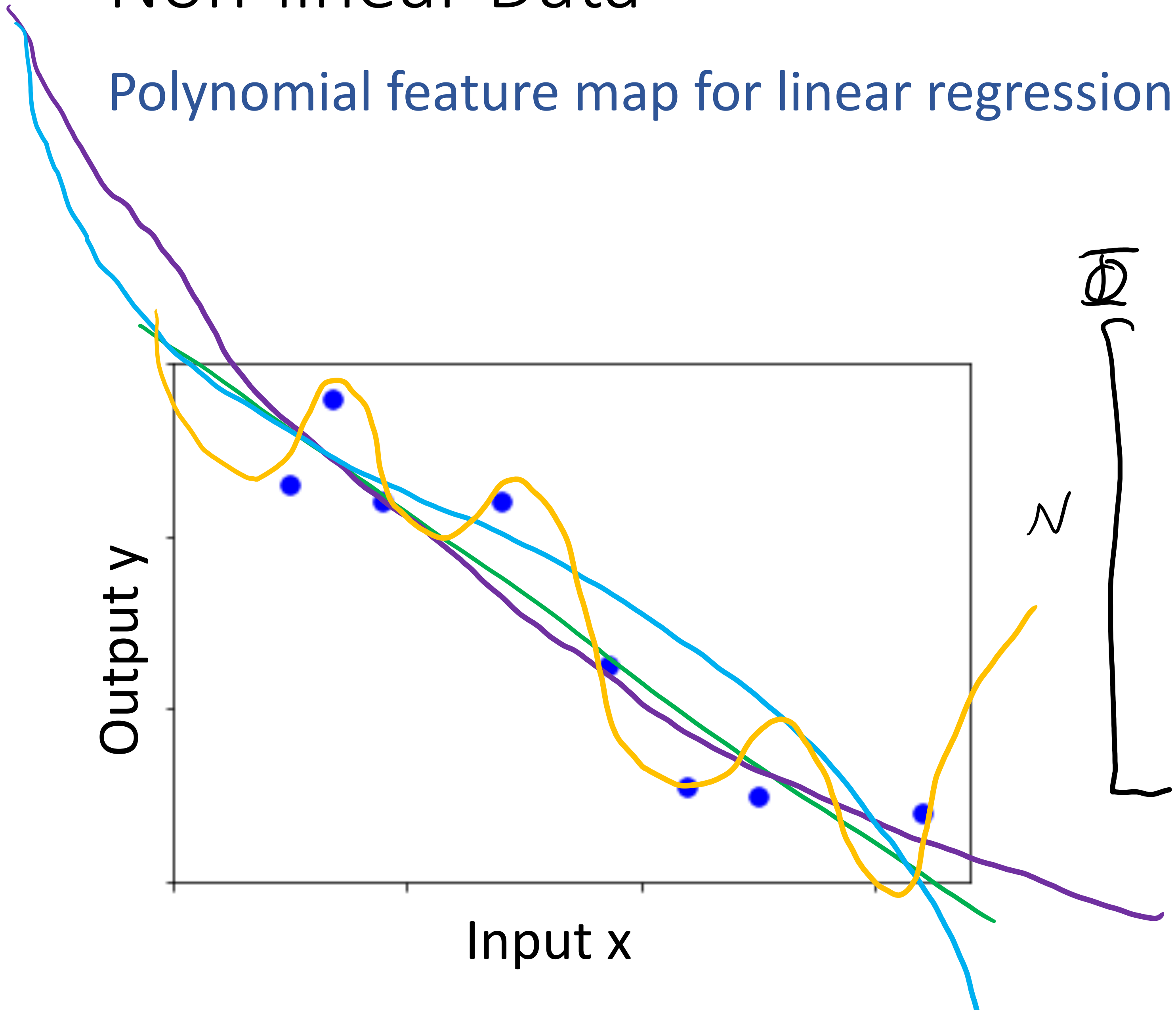
y	x_1	x_2
44	-4	16
27	-3	9
5	-1	1
0	0	0
2	2	4
5	2.5	6.25
9	3	9
35	5	25

Non-linear Data

$$x \longrightarrow \phi(x)$$

$$X \longrightarrow \Phi$$

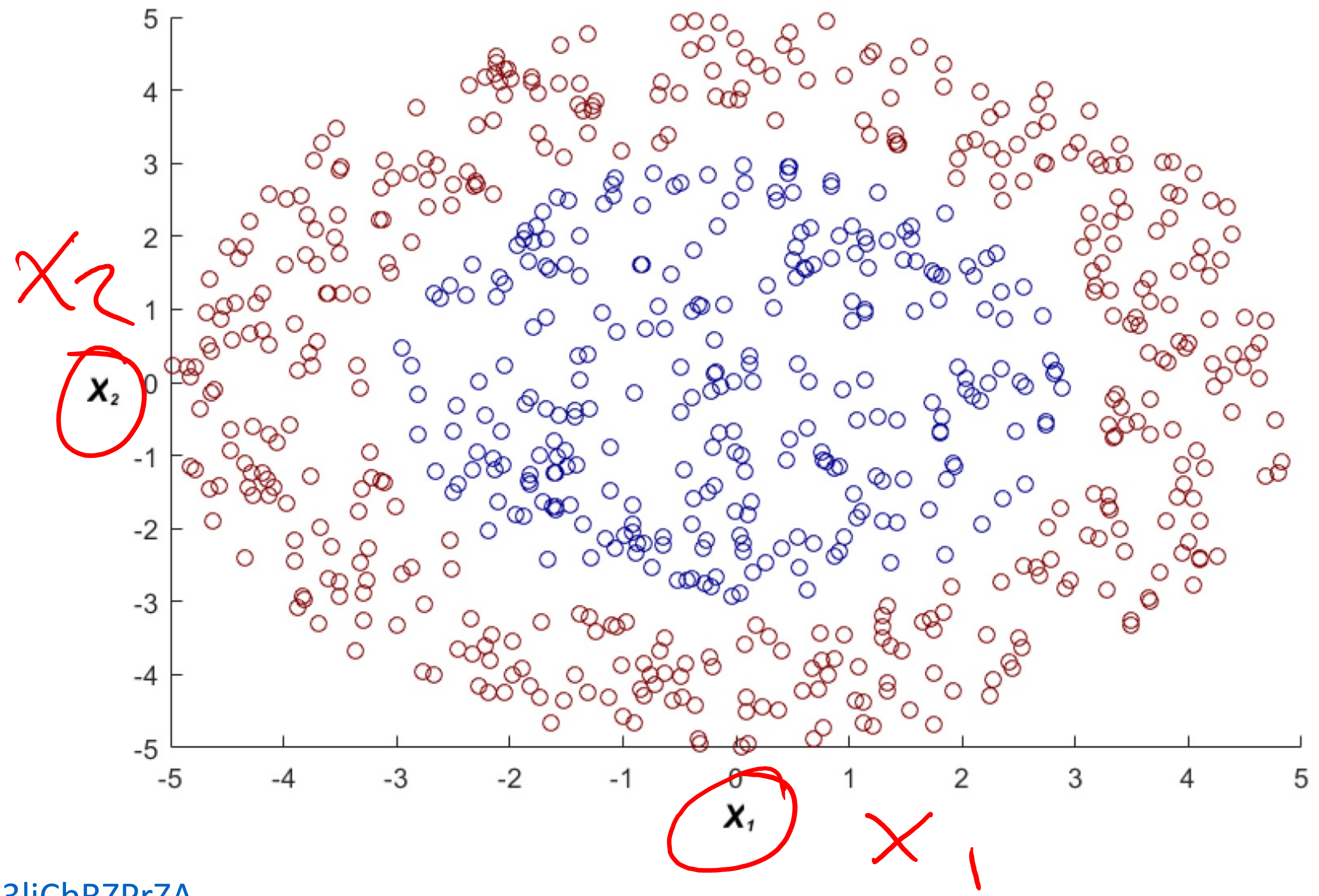
Polynomial feature map for linear regression



$$\Phi = N \left[\begin{array}{ccccc} x^{(1)} & x^{(1)^2} & x^{(1)^3} & \cdots & x^{(1)^{15}} \\ x^{(2)} & x^{(2)^2} & x^{(2)^3} & & \\ \vdots & & & & \\ x^{(N)} & x^{(N)^2} & & & \end{array} \right]$$

Non-linear Data

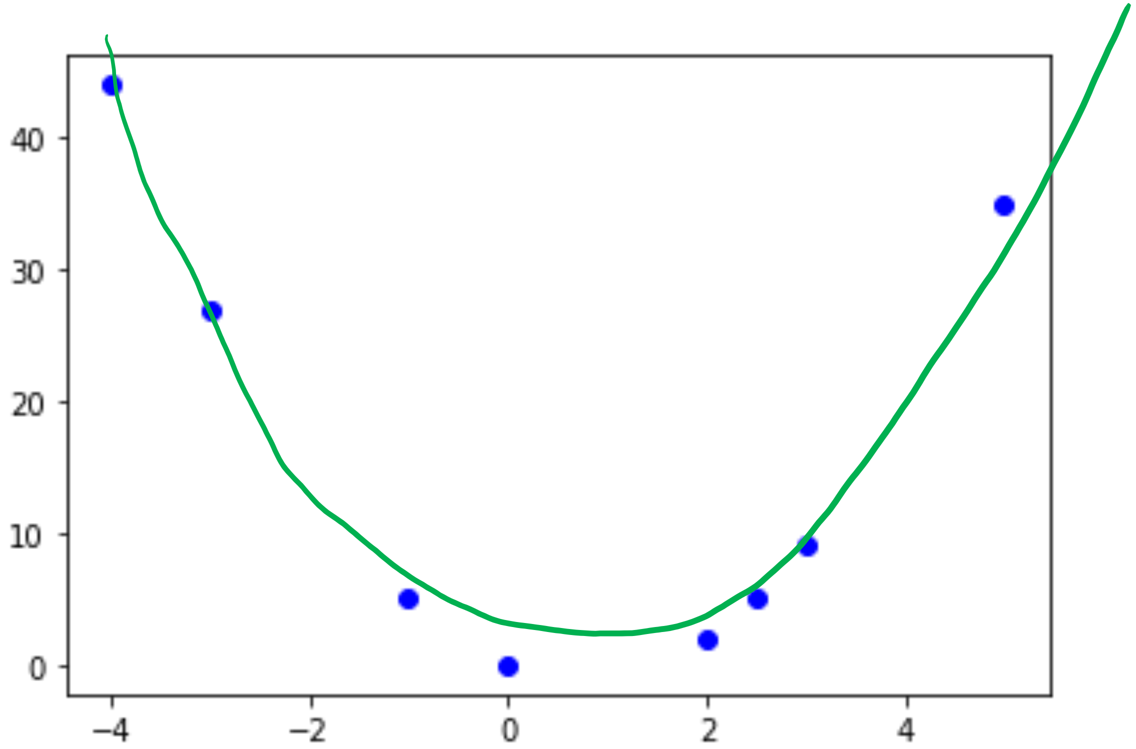
Polynomial feature map for linear classification

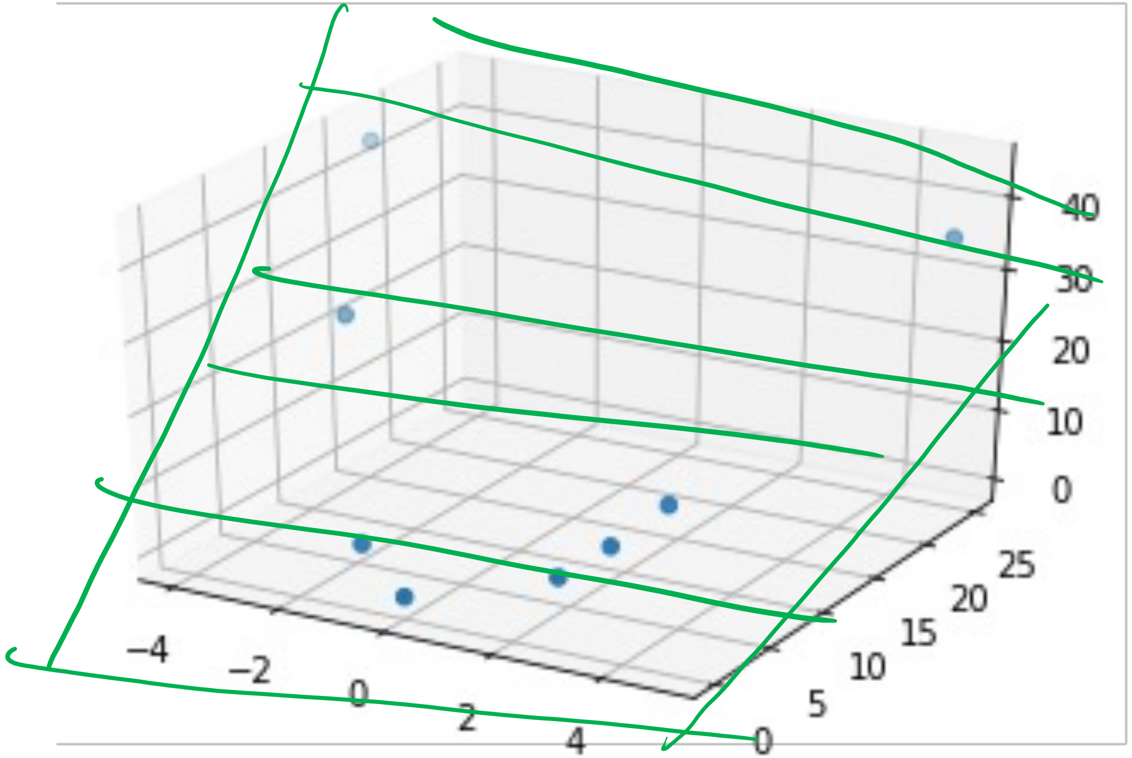


https://www.youtube.com/watch?v=3liCbRZPrZA

Feature Maps

Two ways to think about it



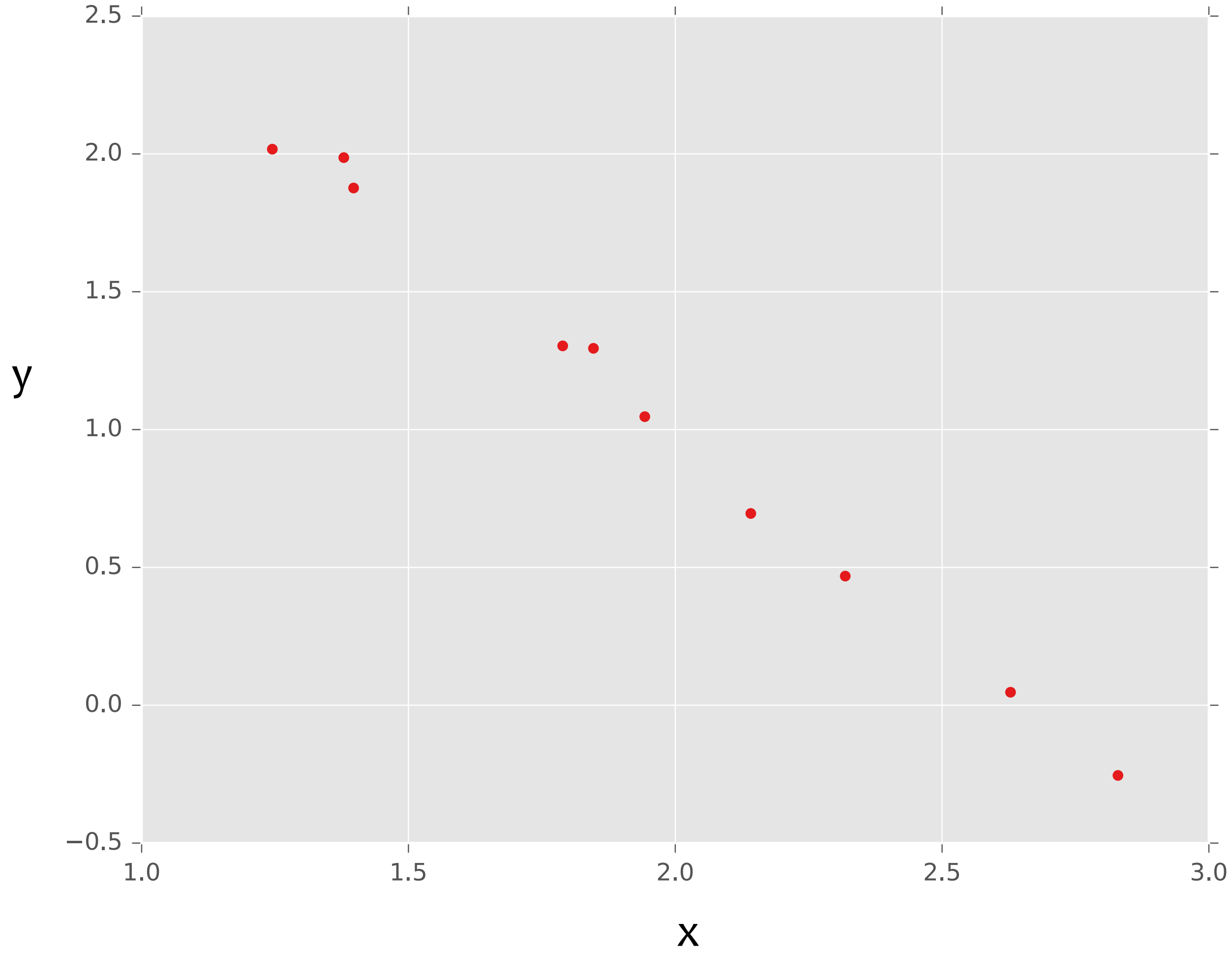


Example: Linear Regression

Goal: Learn y = $\mathbf{w}^T f(\mathbf{x}) + b$ where f(.) is a polynomial basis function

y	x
2.0	1.2
1.3	1.7
0.1	2.7
1.1	1.9

true "unknown" target function is linear with negative slope and gaussian noise



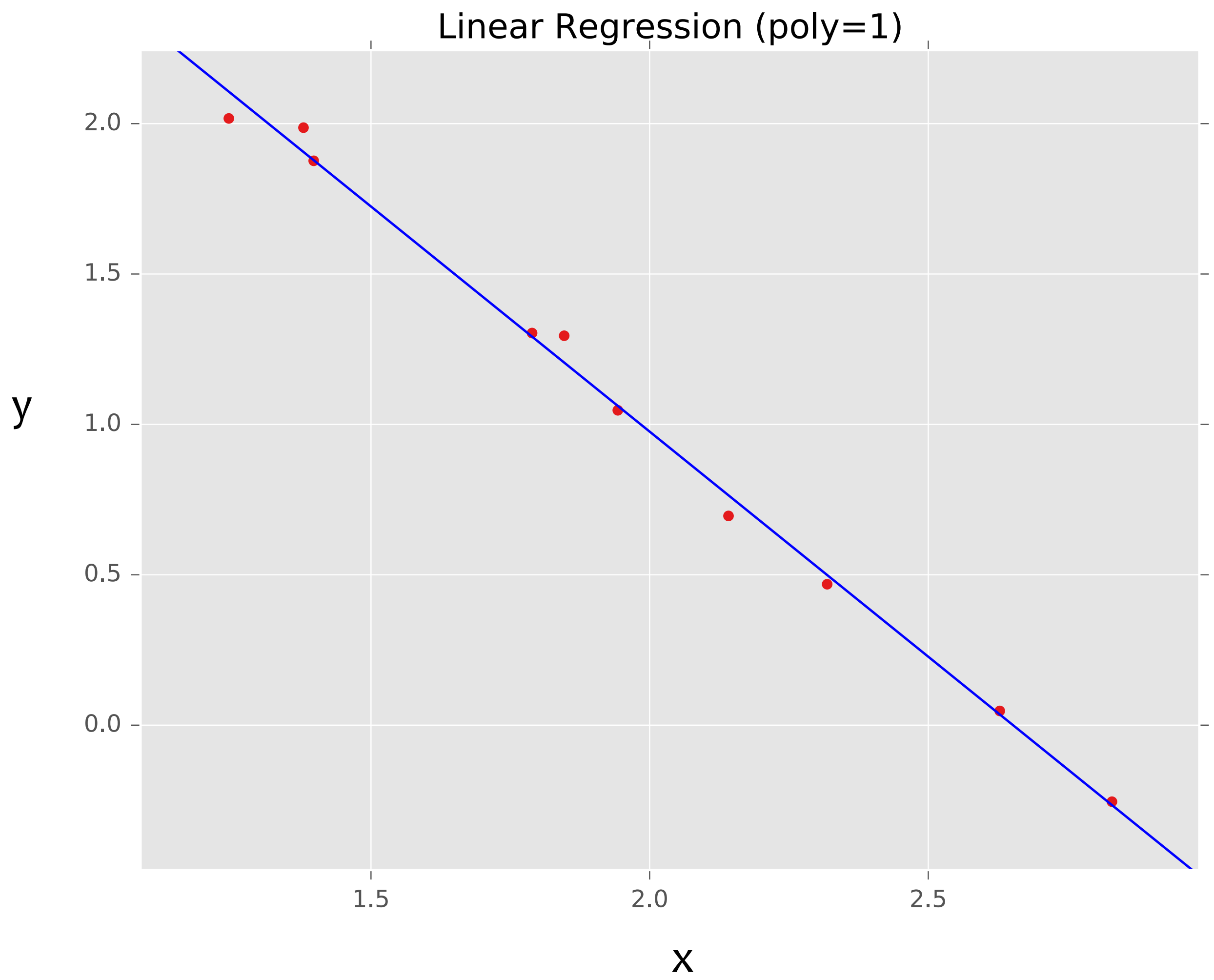
Slide credit: CMU MLD Matt Gormley

Example: Linear Regression

Goal: Learn $y = \mathbf{w}^T f(\mathbf{x}) + b$ where $f(.)$ is a polynomial basis function

y	x
2.0	1.2
1.3	1.7
0.1	2.7
1.1	1.9

true "unknown" target function is linear with negative slope and gaussian noise



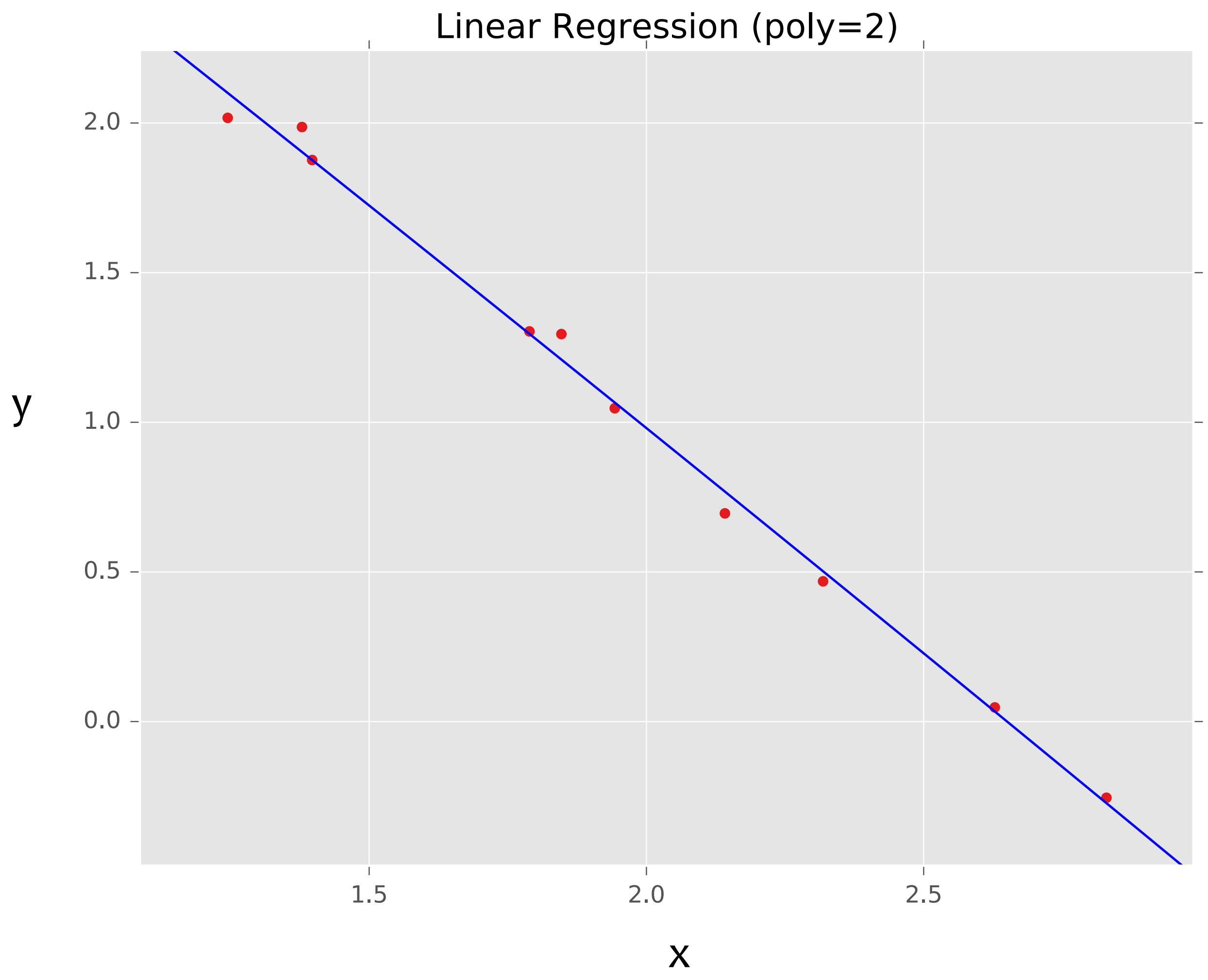
Slide credit: CMU MLD Matt Gormley

Example: Linear Regression

Goal: Learn $y = \mathbf{w}^T f(\mathbf{x}) + b$
where f(.) is a polynomial basis function

y	x	x^2
2.0	1.2	$(1.2)^2$
1.3	1.7	$(1.7)^2$
0.1	2.7	$(2.7)^2$
1.1	1.9	$(1.9)^2$

true "unknown" target function is linear with negative slope and gaussian noise



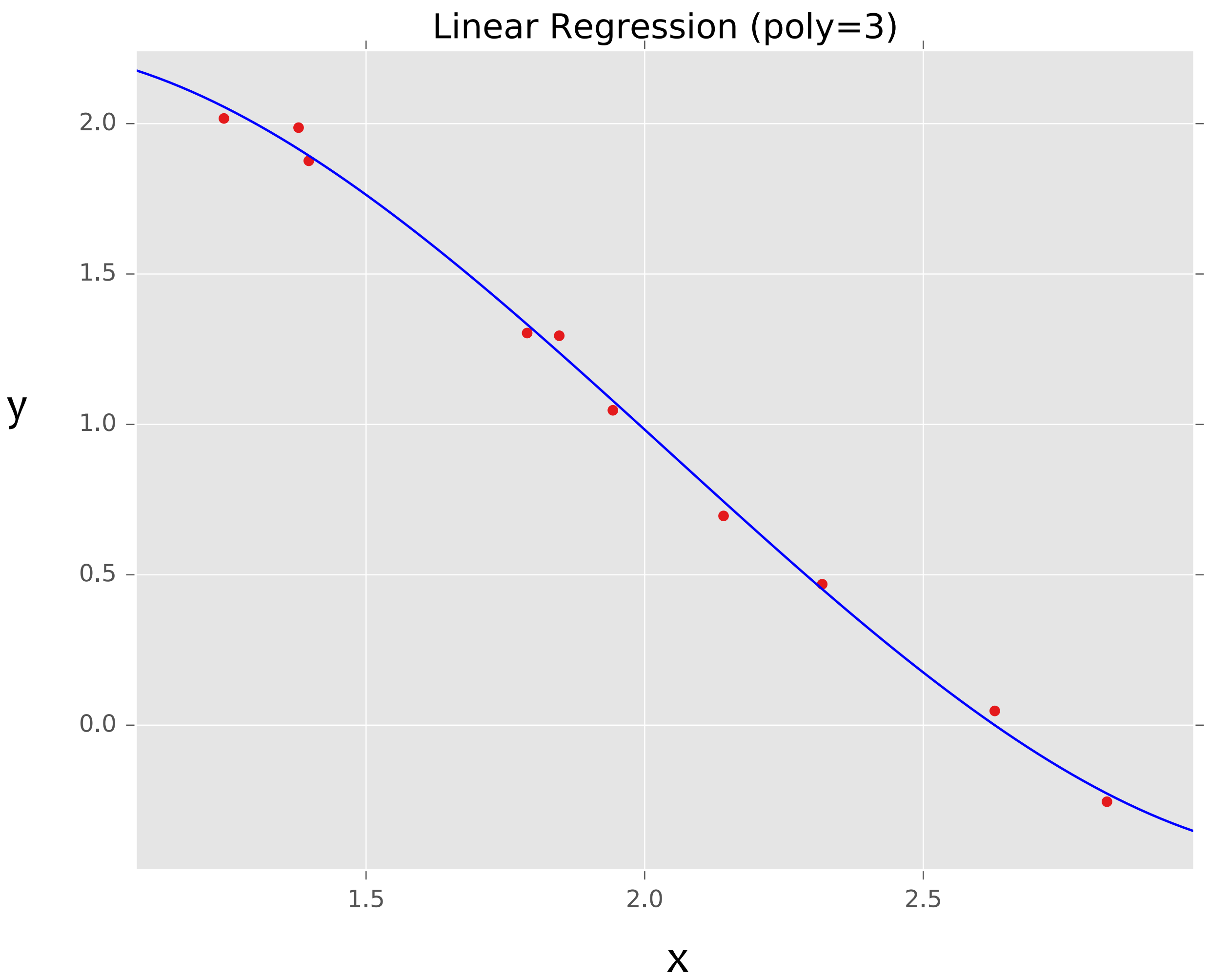
Slide credit: CMU MLD Matt Gormley

Example: Linear Regression

Goal: Learn y = $\mathbf{w}^T f(\mathbf{x}) + b$
where f(.) is a polynomial basis function

y	x	x^2	x^3
2.0	1.2	$(1.2)^2$	$(1.2)^3$
1.3	1.7	$(1.7)^2$	$(1.7)^3$
0.1	2.7	$(2.7)^2$	$(2.7)^3$
1.1	1.9	$(1.9)^2$	$(1.9)^3$

true "unknown" target function is linear with negative slope and gaussian noise

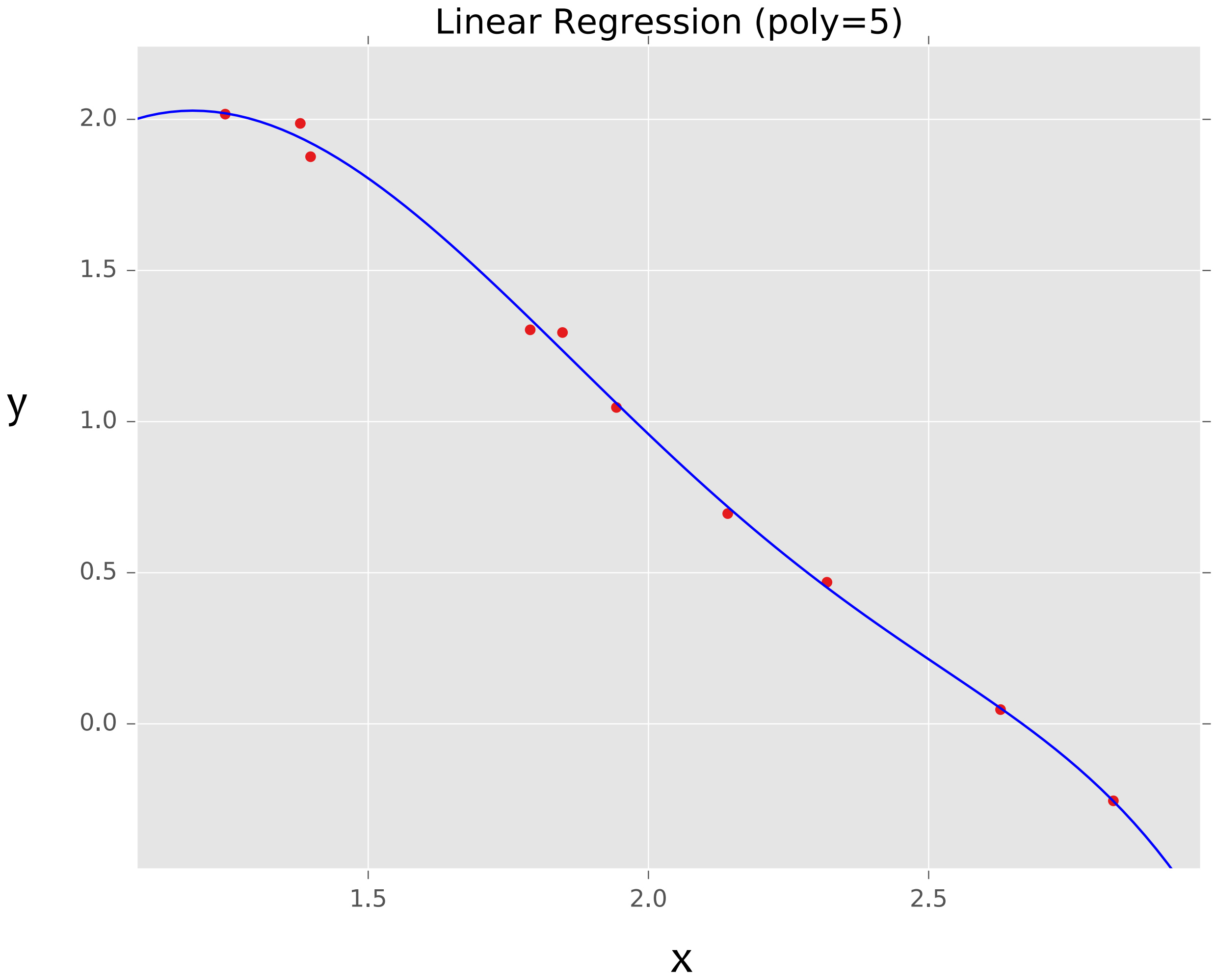


Example: Linear Regression

Goal: Learn y = $\mathbf{w}^T$ f($\mathbf{x}$) + b
where f(.) is a polynomial basis function

y	x	x^2	…	x^5
2.0	1.2	$(1.2)^2$	…	$(1.2)^5$
1.3	1.7	$(1.7)^2$	…	$(1.7)^5$
0.1	2.7	$(2.7)^2$	…	$(2.7)^5$
1.1	1.9	$(1.9)^2$	…	$(1.9)^5$

true "unknown" target function is linear with negative slope and gaussian noise



Slide credit: CMU MLD Matt Gormley

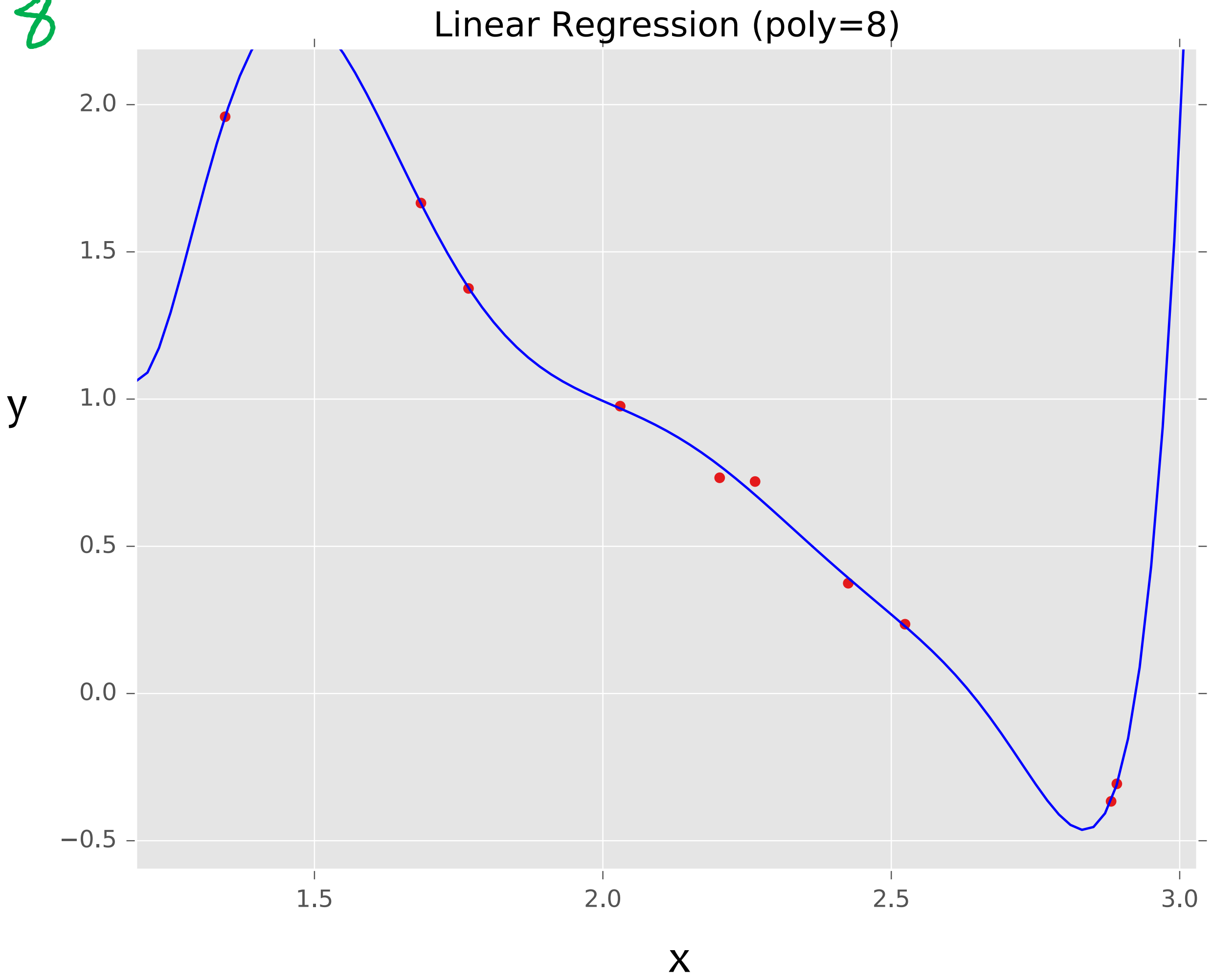
Example: Linear Regression

Goal: Learn y = $\mathbf{w}^T$ f($\mathbf{x}$) + b
where f(.) is a polynomial basis function

x^8

y	x	x^2	…	x^8
2.0	1.2	$(1.2)^2$	…	$(1.2)^8$
1.3	1.7	$(1.7)^2$	…	$(1.7)^8$
0.1	2.7	$(2.7)^2$	…	$(2.7)^8$
1.1	1.9	$(1.9)^2$	…	$(1.9)^8$

true "unknown" target function is linear with negative slope and gaussian noise



Slide credit: CMU MLD Matt Gormley

Example: Linear Regression

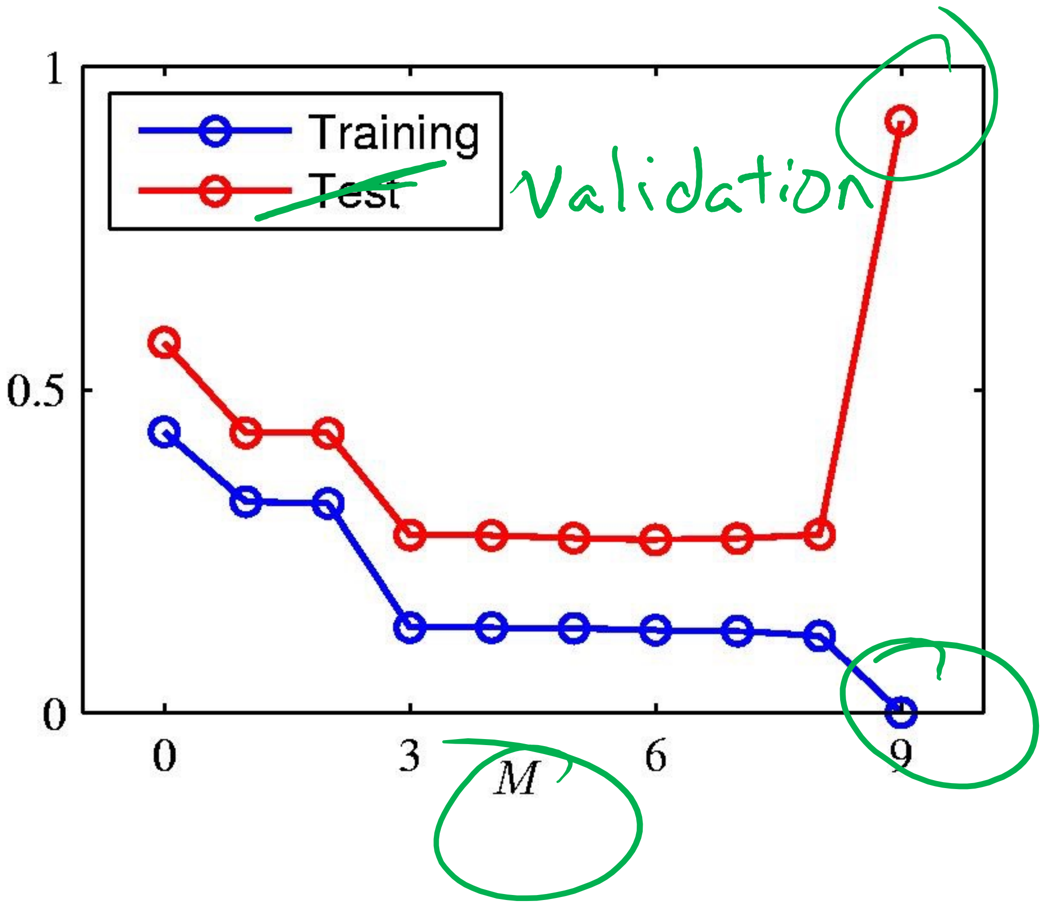
Goal: Learn $y = \mathbf{w}^T f(\mathbf{x}) + b$
where f(.) is a polynomial
basis function

y	x	x^2	…	x^9
2.0	1.2	$(1.2)^2$	…	$(1.2)^9$
1.3	1.7	$(1.7)^2$	…	$(1.7)^9$
0.1	2.7	$(2.7)^2$	…	$(2.7)^9$
1.1	1.9	$(1.9)^2$	…	$(1.9)^9$

true "unknown" target function is linear with negative slope and gaussian noise

Slide credit: CMU MLD Matt Gormley

Over-fitting



Slide credit: CMU MLD William Cohen

Polynomial Coefficients

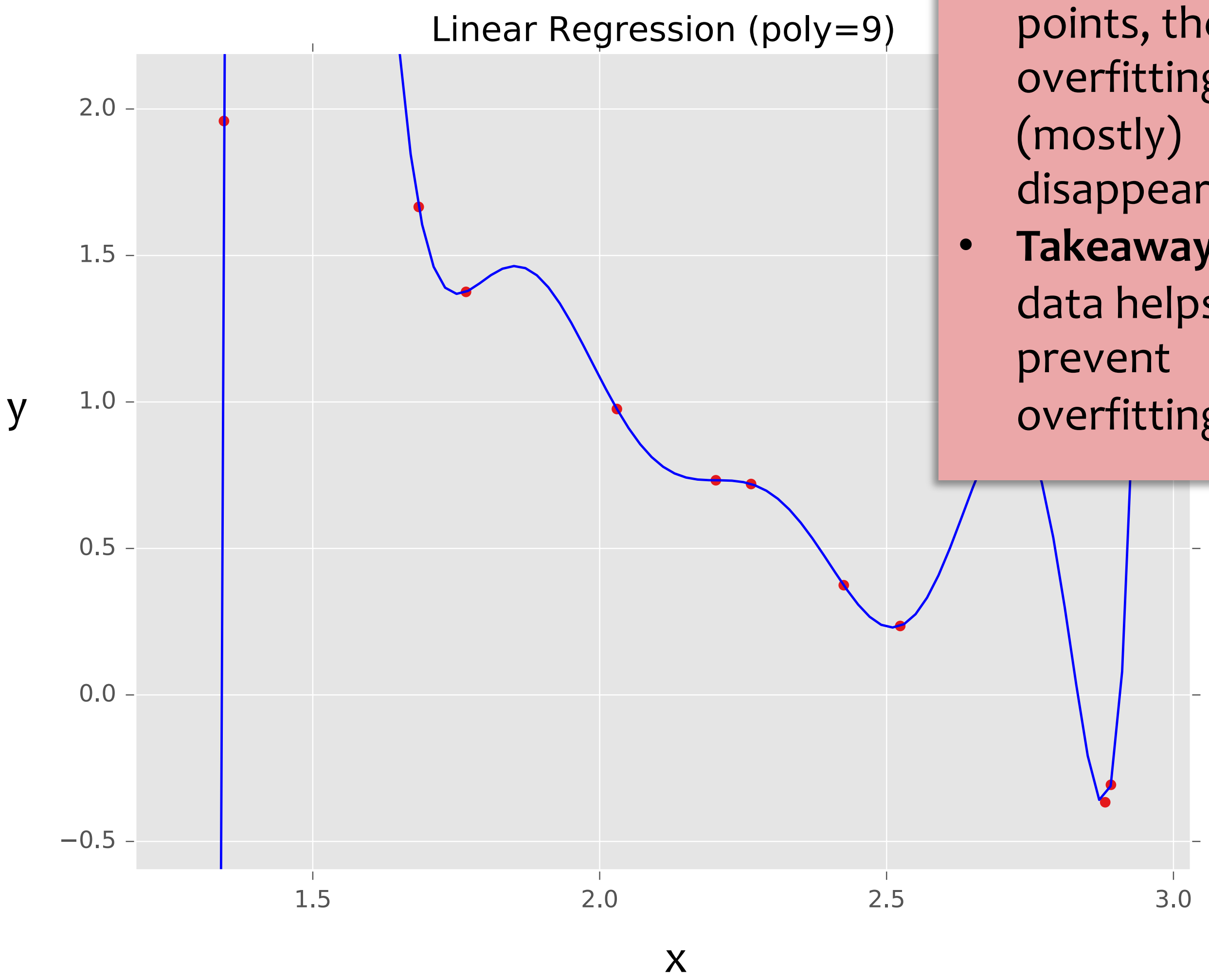
	$M=0$	$M=1$	$M=3$	$M=9$
θ_0	0.19	0.82	0.31	0.35
θ_1		-1.27	7.99	232.37
θ_2			-25.43	-5321.83
θ_3			17.37	48568.31
θ_4				-231639.30
θ_5				640042.26
θ_6				-1061800.52
θ_7				1042400.18
θ_8				-557682.99
θ_9				125201.43

Slide credit: CMU MLD William Cohen

Example: Linear Regression

Goal: Learn $y = \mathbf{w}^T f(\mathbf{x}) + b$ where f(.) is a polynomial basis function

i	y	x	…	x^9
1	2.0	1.2	…	$(1.2)^9$
2	1.3	1.7	…	$(1.7)^9$
…	…	…	…	…
10	1.1	1.9	…	$(1.9)^9$



- With just N = 10 points we overfit!
- But with N = 100 points, the overfitting (mostly) disappears
- **Takeaway:** more data helps prevent overfitting

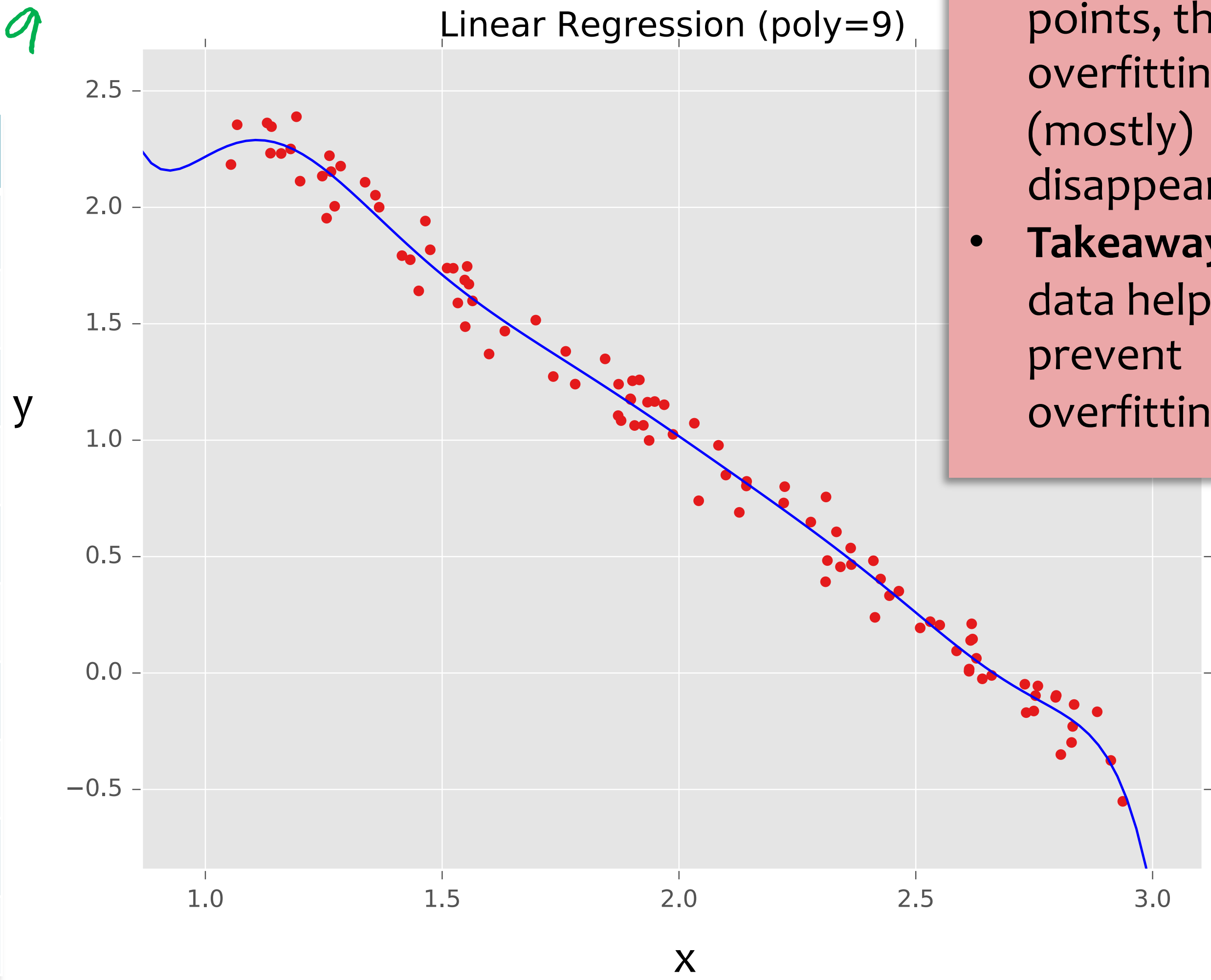
Slide credit: CMU MLD Matt Gormley

Example: Linear Regression

Goal: Learn y = $\mathbf{w}^T f(\mathbf{x}) + b$ where f(.) is a polynomial basis function

x^9

i	y	x	…	x^9
1	2.0	1.2	…	$(1.2)^9$
2	1.3	1.7	…	$(1.7)^9$
3	0.1	2.7	…	$(2.7)^9$
4	1.1	1.9	…	$(1.9)^9$
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
98	…	…	…	…
99	…	…	…	…
100	0.9	1.5	…	$(1.5)^9$



- With just N = 10 points we overfit!
- But with N = 100 points, the overfitting (mostly) disappears
- **Takeaway:** more data helps prevent overfitting

Slide credit: CMU MLD Matt Gormley

How Do We Deal with Real-world Problems

Politician Voting Classification

y	$\phi_1(x)$	$\phi_2(x)$
0	1	0
1	1	1
1	0	1

Party	vote_1	vote_2
y	x_1	x_2
Dem	yes	no
Rep	yes	yes
Rep	no	yes

$\phi(\vec{x}_{raw}) \rightarrow \overrightarrow{\text{feature}}$

How Do We Deal with Real-world Problems

SPAM Classification

$X = \{ \underline{\text{email body}}, \underline{\text{subject}}, \underline{\text{from}} \}$

$\downarrow$

$\phi(X)$

$$\begin{bmatrix} \phi_1(x) \\ \phi_2(x) \\ \\ \\ \end{bmatrix} \begin{matrix} \longleftarrow \text{\# of all caps words} \\ \longleftarrow \text{0/1 has my name at top} \\ \\ \\ \end{matrix}$$

How Do We Deal with Real-world Problems

SPAM Classification

$X = \{ \text{email body}, \text{subject}, \text{from} \}$

$\phi(X)$

$$\begin{bmatrix} \phi_1(X) \\ \phi_2(X) \\ \\ \\ \end{bmatrix}$$

Vocab

1 a
2 aardv
3 apple
⋮

How Do We Deal with Real-world Problems

Predicting Rating from Written Movie Review

$$x = \{\text{text}\}$$

$$\phi(x) \in \mathbb{R}^m$$

$$\begin{bmatrix} \phi_1(x) \\ \phi_2(x) \\ \vdots \\ \phi_m(x) \end{bmatrix}$$

$\phi_1(x) \leftarrow$ # times fantastic

$$\hat{y} = h(\vec{x})$$

$$= g(\vec{\theta}^T \vec{x})$$

$$\downarrow$$

$$= g(\vec{\theta}^T \phi(x))$$

How Do We Deal with Real-world Problems

Images: Handwritten Digits

$$\vec{x} \in \mathbb{R}^{784}$$

$$g(\vec{\theta}^T \vec{x})$$

$$\begin{bmatrix} \phi_1(\vec{x}) \\ \phi_2(\vec{x}) \\ \\ \phi_n(\vec{x}) \end{bmatrix} \leftarrow \text{\# holes}$$

28

28

How Do We Deal with Real-world Problems

Images: Face Recognition

How Do We Deal with Real-world Problems

Images: Animal Classification

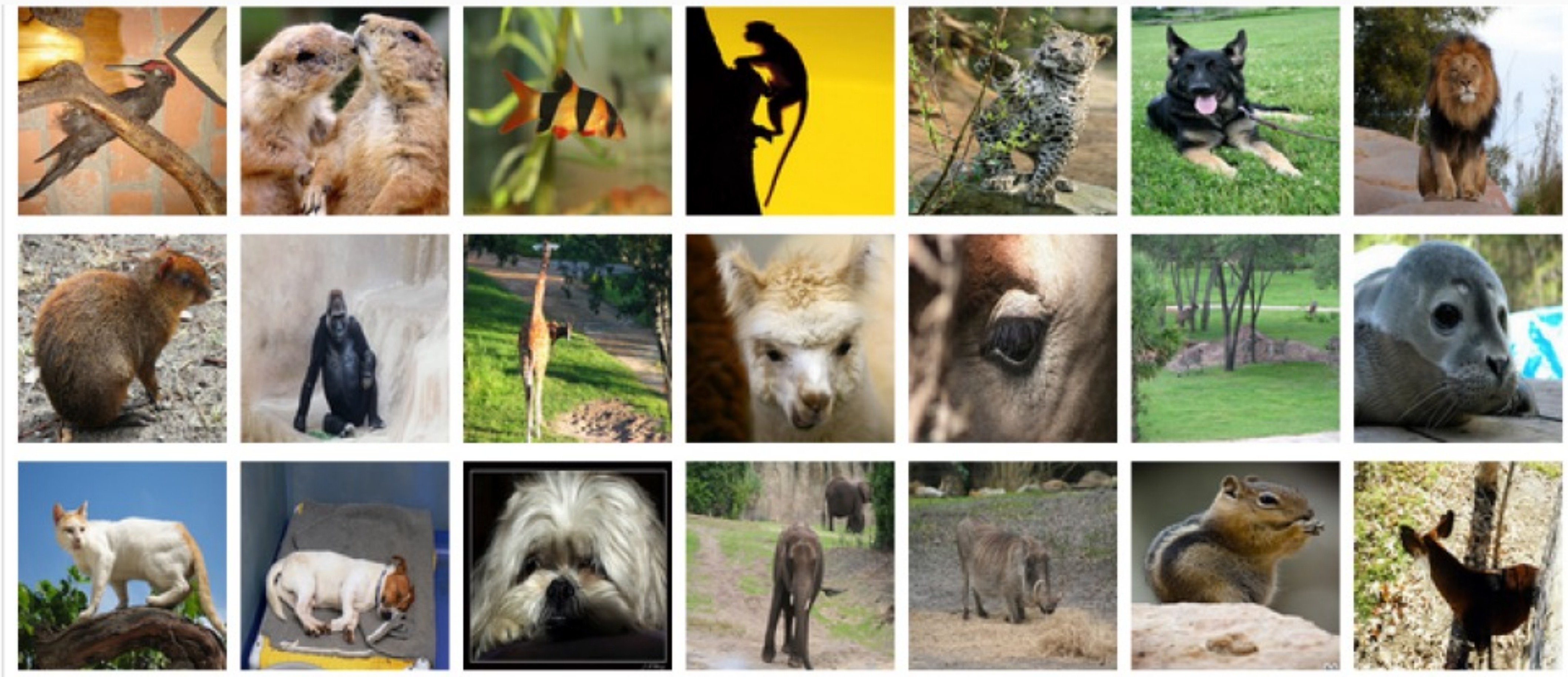


Image: ImageNet

How Do We Deal with Real-world Problems

Images: Animal Classification

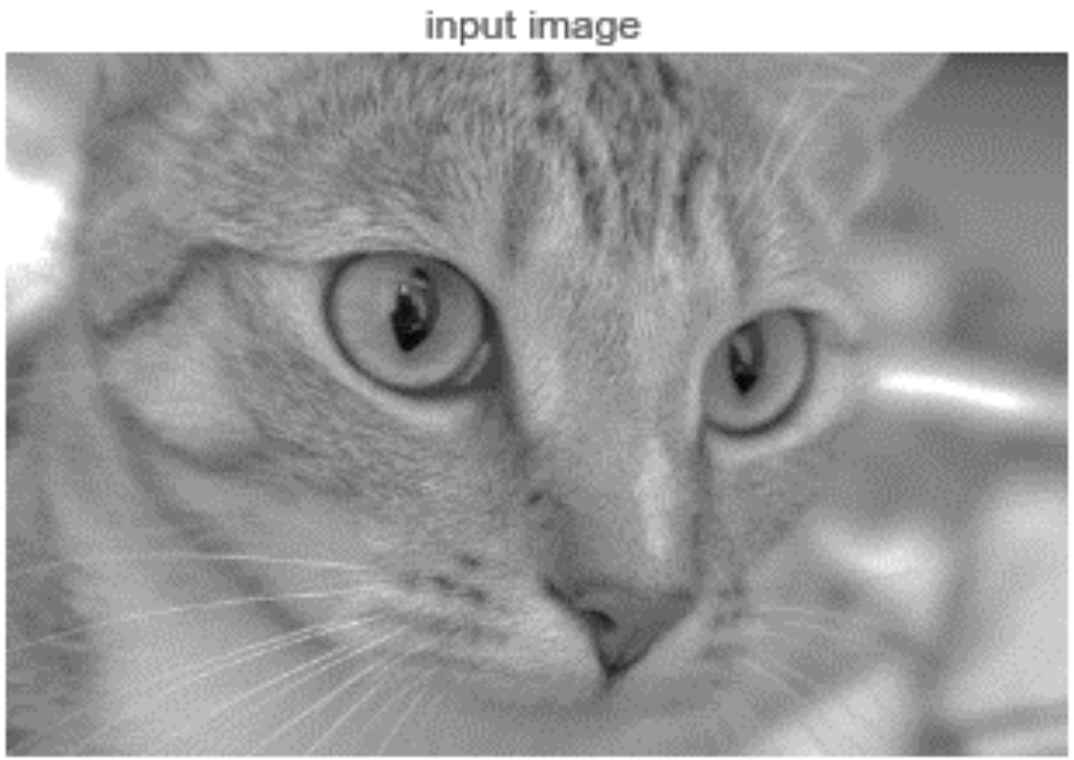


Image: https://jakevdp.github.io/PythonDataScienceHandbook/05.14-image-features.html

How Do We Deal with Real-world Problems

Preview: Neural Networks

input image

visualization of HOG features

Image: https://jakevdp.github.io/PythonDataScienceHandbook/05.14-image-features.html

How Do We Deal with Real-world Problems

Preview: Neural Networks

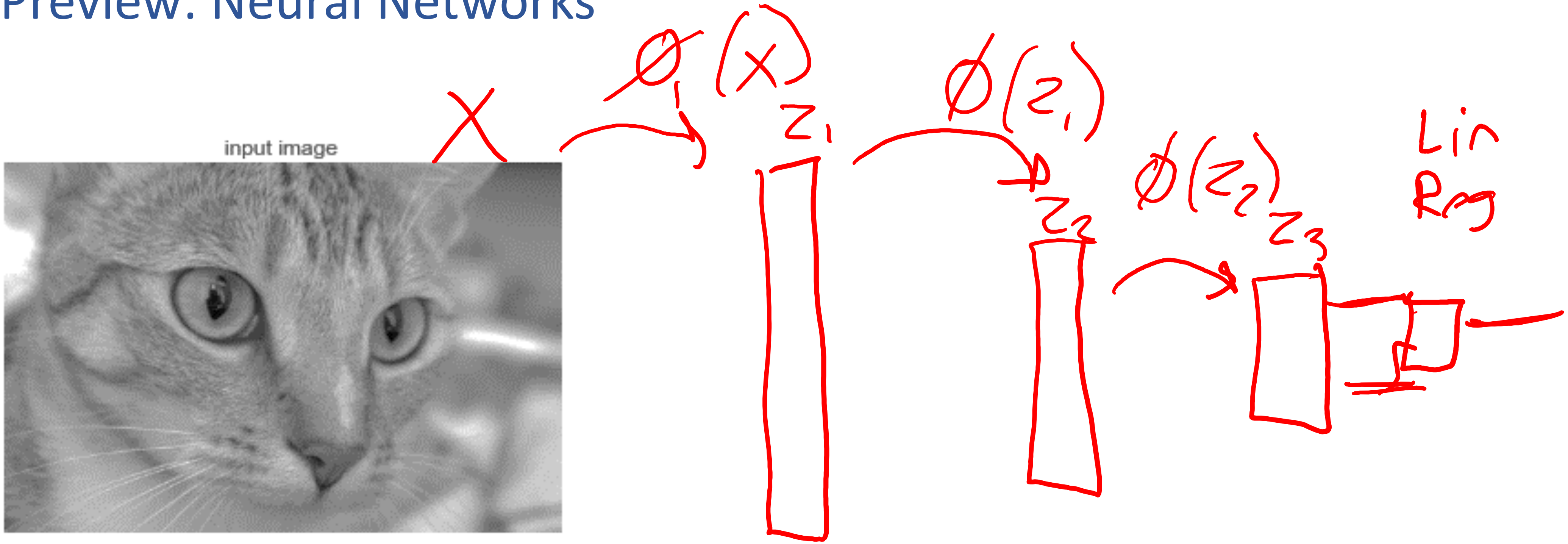


Image: https://jakevdp.github.io/PythonDataScienceHandbook/05.14-image-features.html