LECTURE 2 TERM 2:

MSIN0097

UCL SCHOOL OF MANAGEMENT

Predictive Analytics

Video 6: Regression

A P MOORE

A. ClAssification

Model requirements

Classifiation
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

B. Regression

Model requirements

Regression
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

Supervised

C. Clustering

Model requirements

Clustering
$x \in [-\infty, \infty]$
$y \in \{0, N\}$

D. Decomposition

Model requirements

Decomposition
$x \in [-\infty, \infty]$
$y \in [-\infty, \infty]$

Unsupervised

END-TO-END

- Discover
- Explore
- Visualize

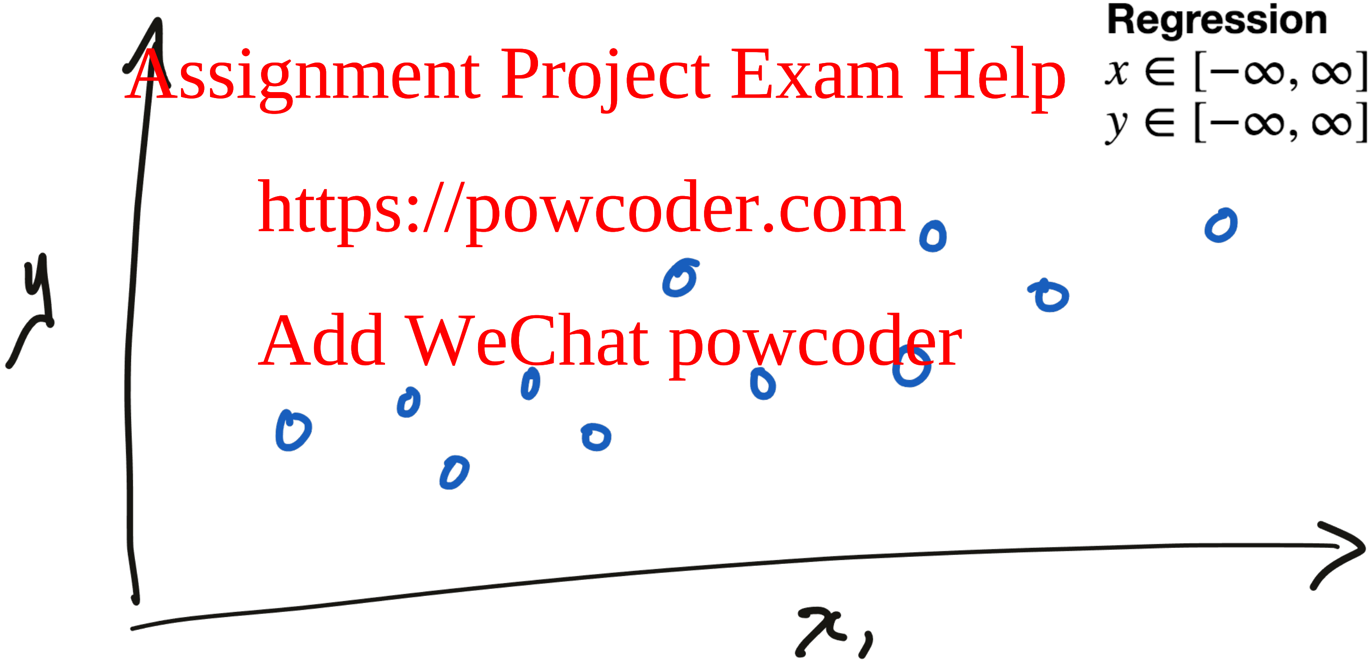
- Clean
- Sample
- Impute
- Encode
- Transform
- Modeling
  - Overfitting
  - Model Selection
    - Learning curves
  - Regularization
    - Degrees of freedom
  - Generalization

- Documentation
- Presentation

- Launch
- Monitor
- Maintain

B. REGRESSION

REAL VALUED VARIABLE



Model requirements

Regression

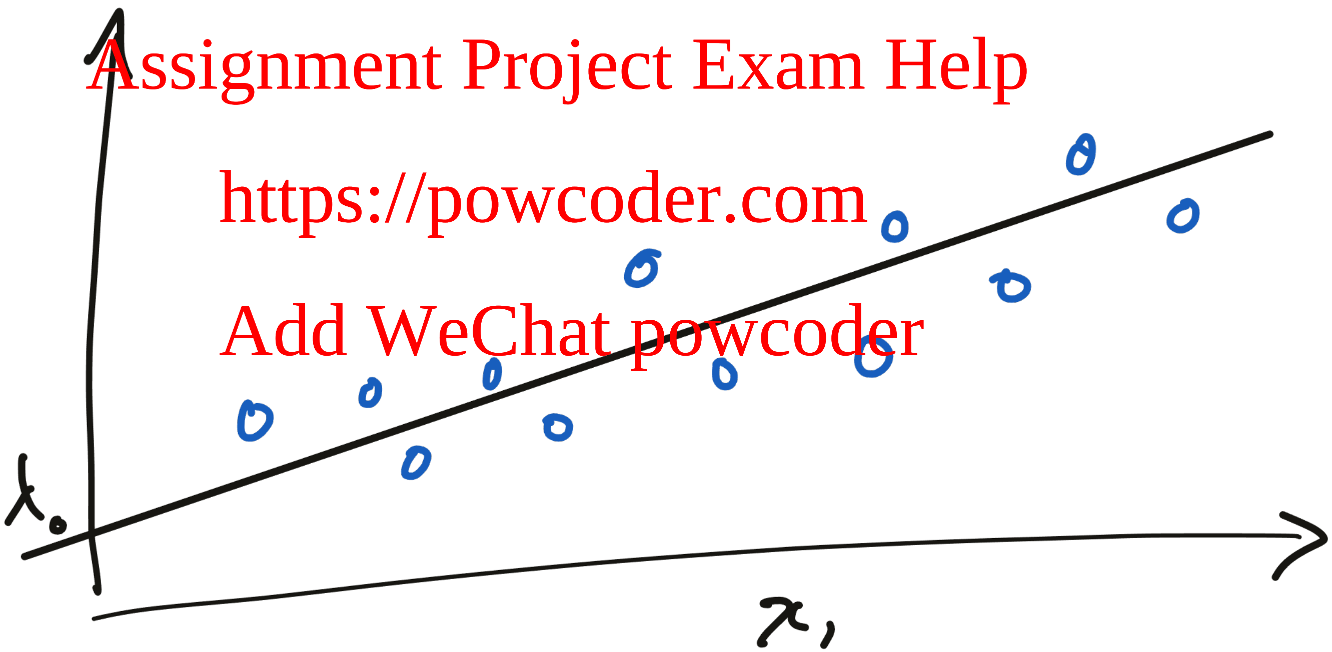
$x \in [-\infty, \infty]$

$y \in [-\infty, \infty]$

LINEAR REGRESSION

REAL VALUED VARIABLE

UCL SCHOOL OF MANAGEMENT



$$\widehat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

LINEAR REGRESSION

Measured data
Features

Inferred/Predicted/Estimated value

True initial value
(world state)

$$x \rightarrow \hat{x} \rightarrow f(x) = \hat{y} \rightarrow y$$

True target value
(world state)

Learned/Fitted function
From n observations

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

predicted value

parameter vector

$$\hat{y} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta}^{\mathsf{T}} \cdot \mathbf{x}$$

hypothesis function

feature vector

COST FUNCTION

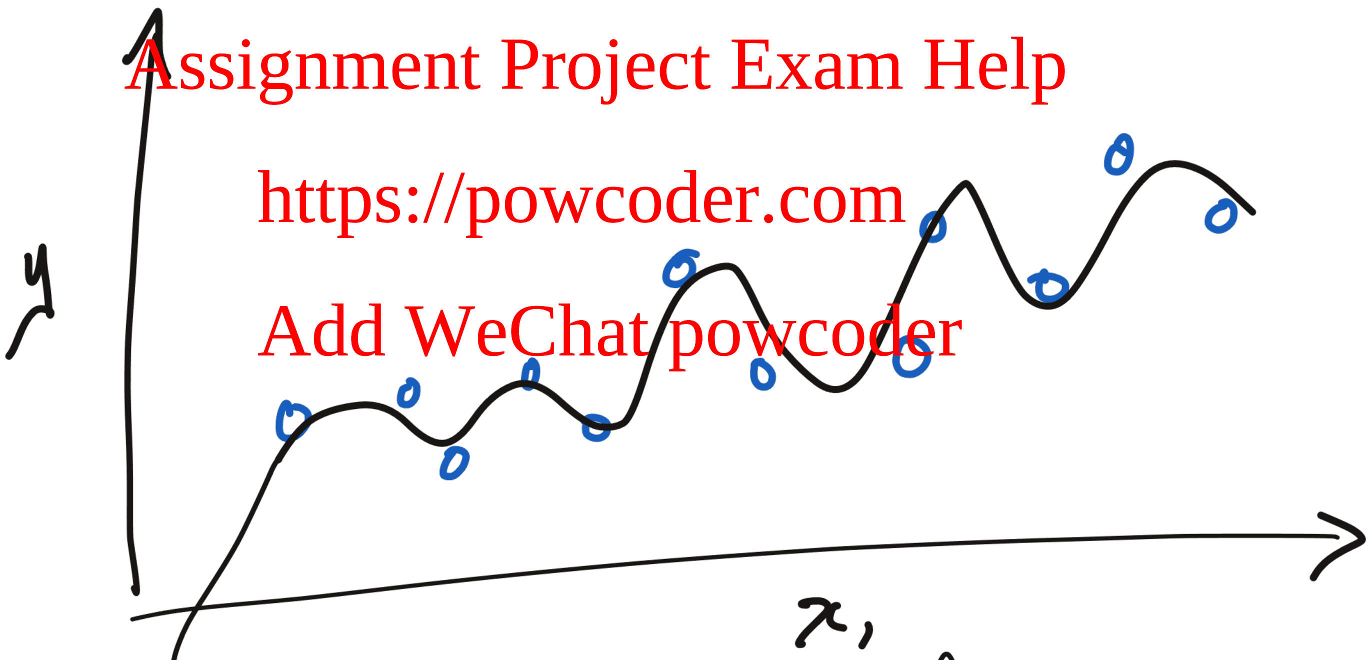
$$\hat{y} = h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta} \cdot \mathbf{x}$$

$$\text{MSE}(\mathbf{X}, h_{\boldsymbol{\theta}}) = \frac{1}{m} \sum_{i=1}^{m} \left(\boldsymbol{\theta}^{\intercal} \mathbf{x}^{(i)} - y^{(i)} \right)^{2}$$

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^{\intercal} \mathbf{X} \right)^{-1} \mathbf{X}^{\intercal} \mathbf{y}$$

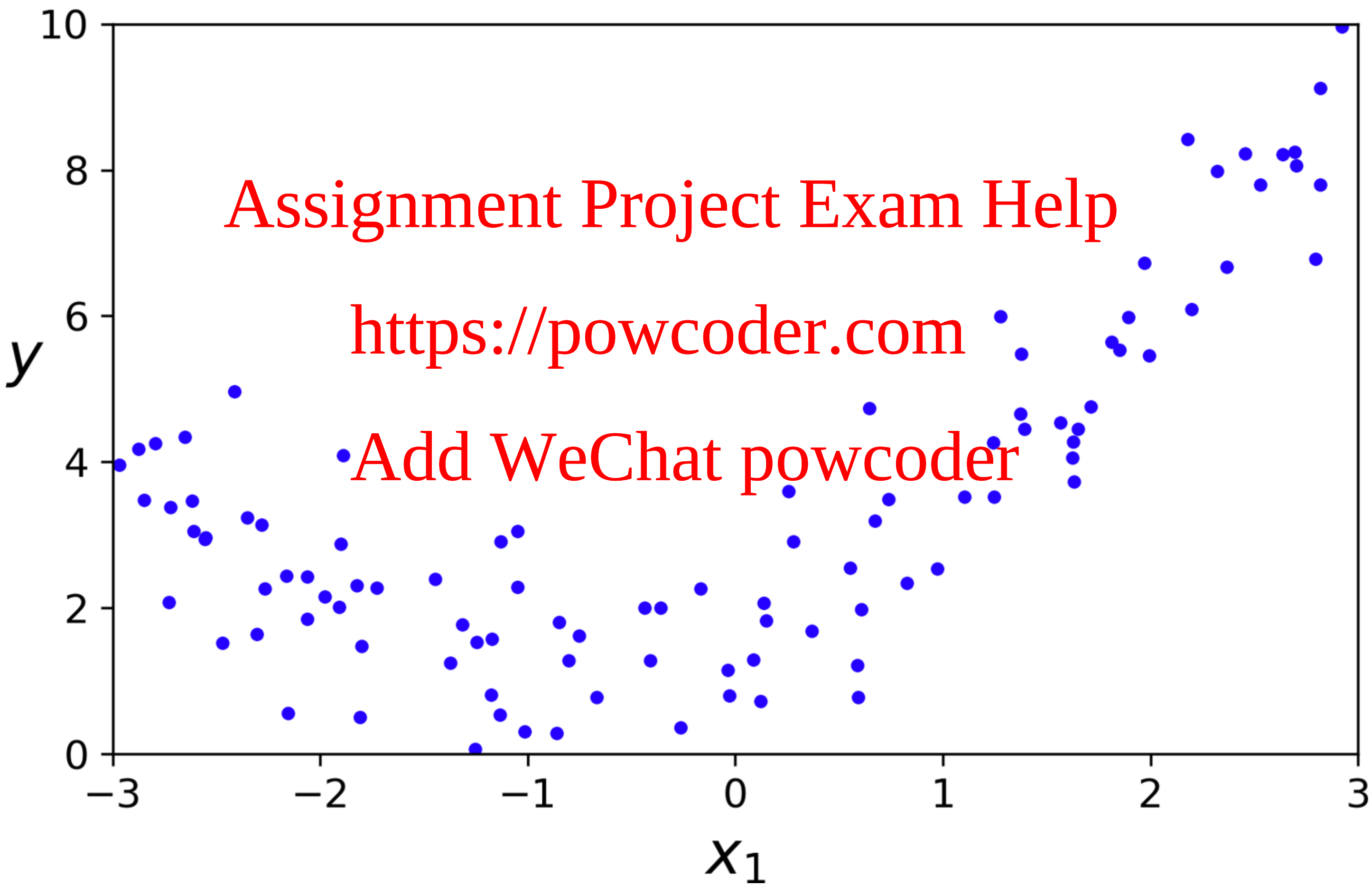
Closed-form solution— Normal Equation

POLYNOMIAL REGRESSION

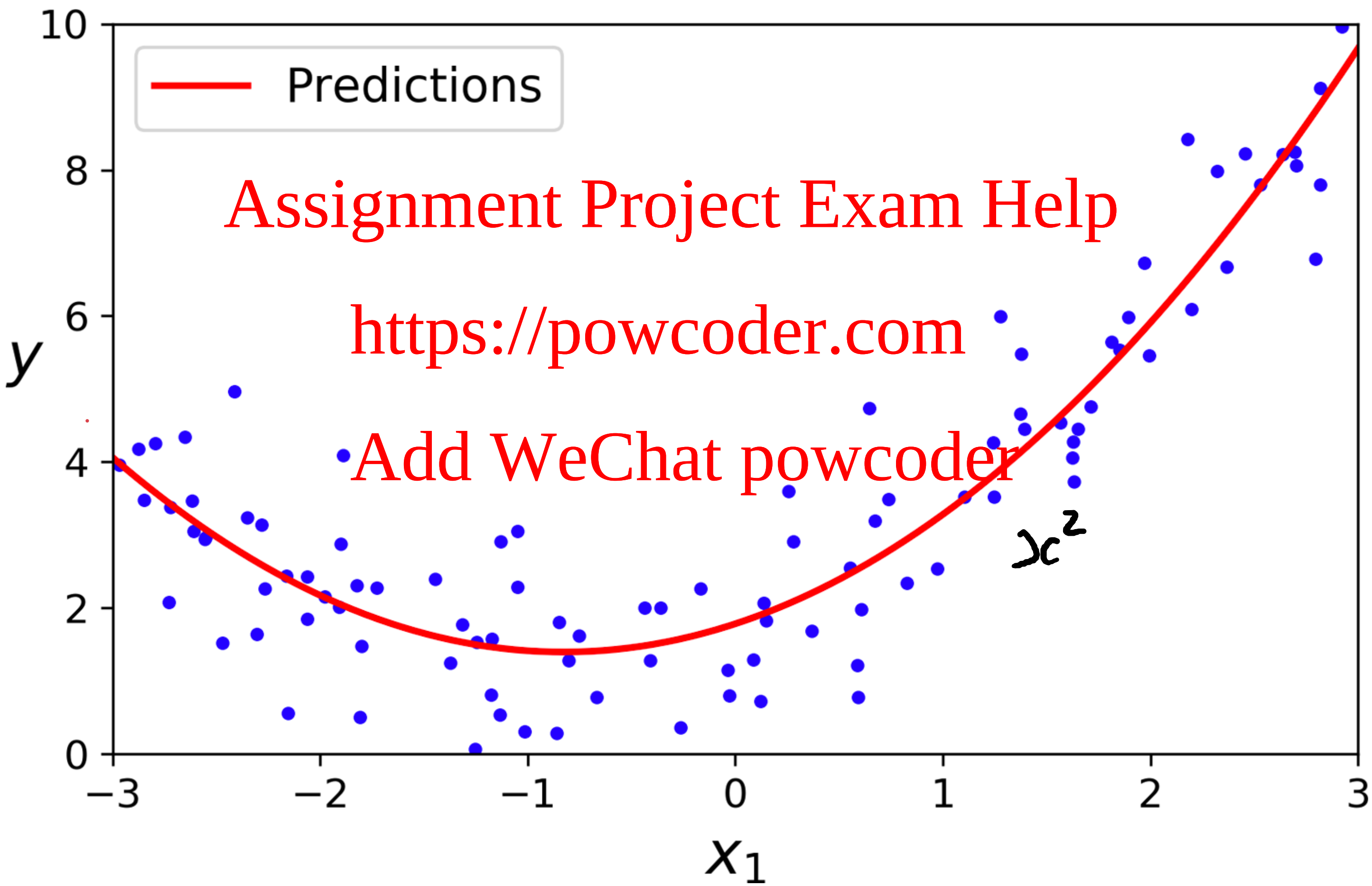


$f(x) = x_1 \;\&\; x_1^2 \;\&\; x_1^3 \ldots$

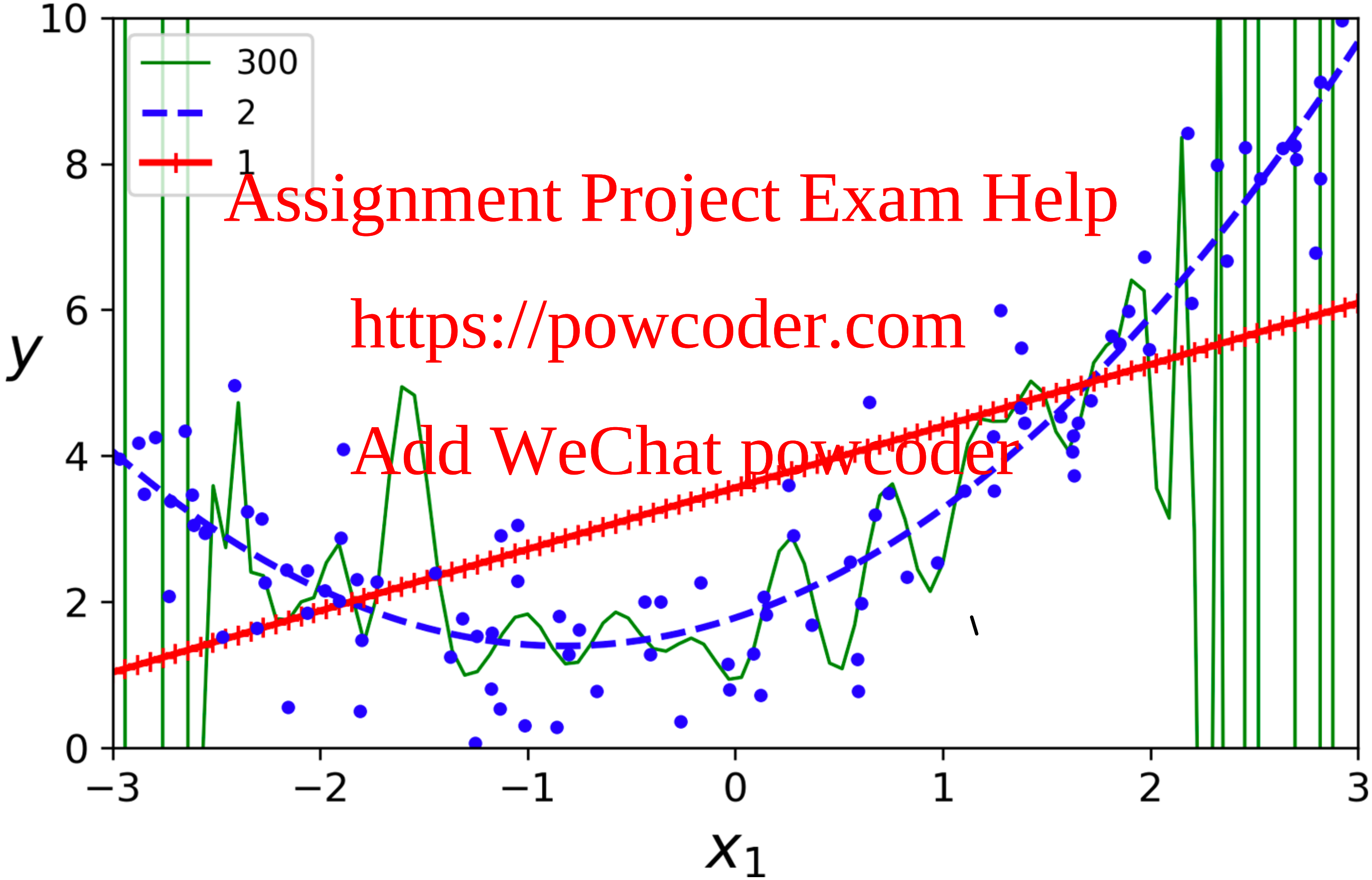
```
m = 100
X = 6 * np.random.rand(m, 1) - 3
y = 0.5 * X**2 + X + 2 + np.random.randn(m, 1)
```



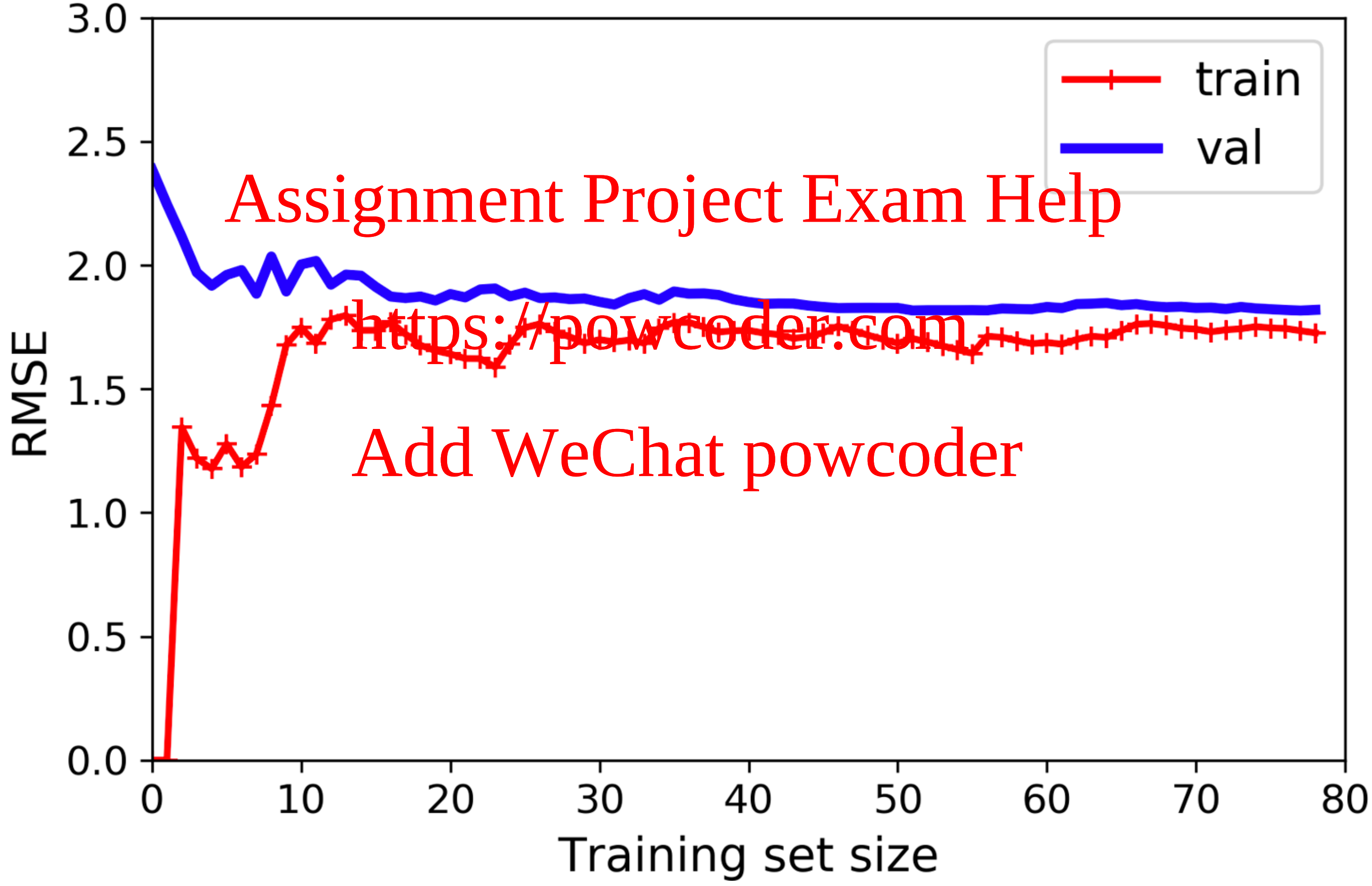
POLYNOMIAL REGRESSION



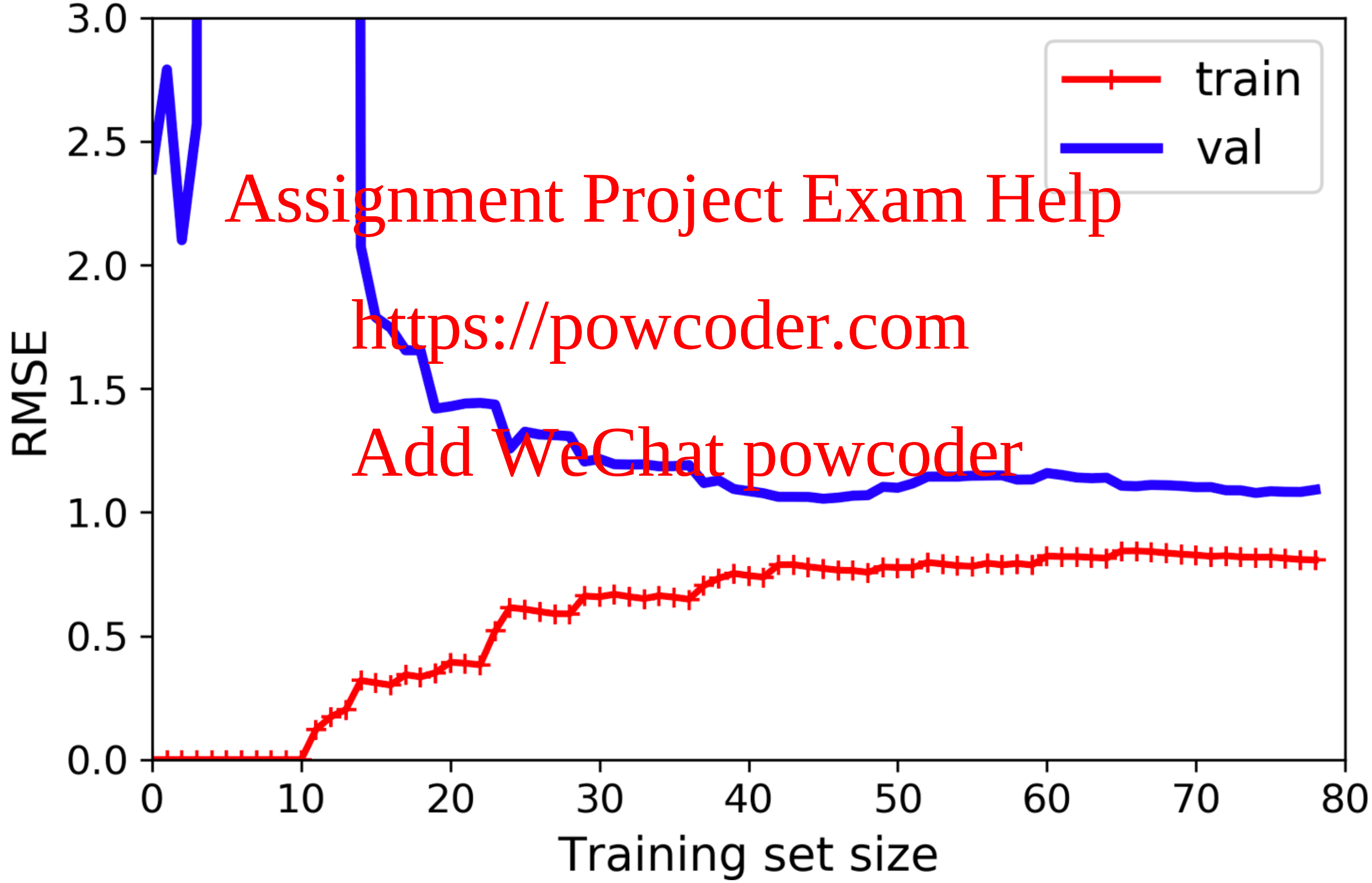
DEGREES OF FREEDOM



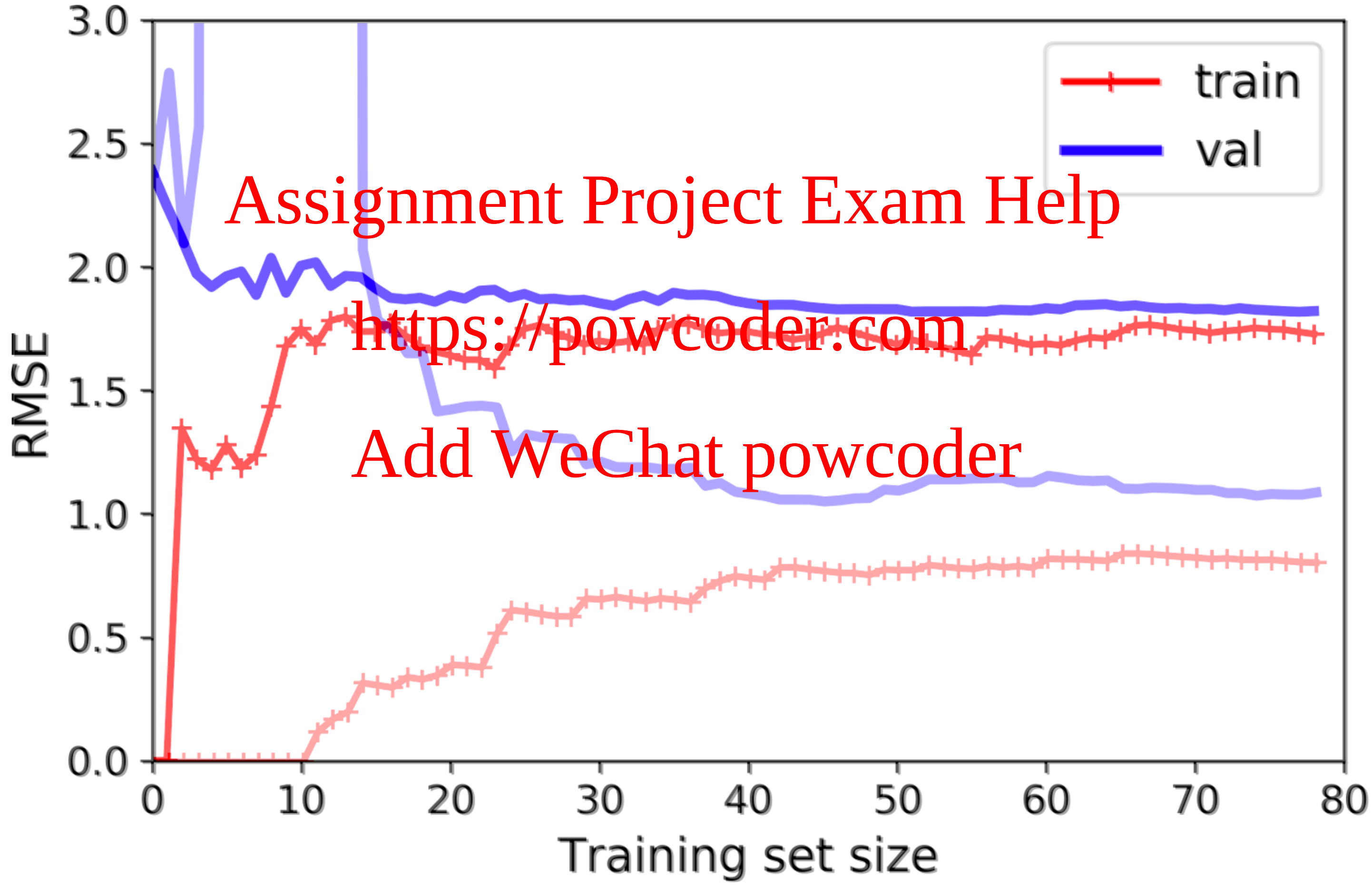
LEARNING CURVES



10^{TH} DEG POLYNOMIAL



LEARNING CURVES



BIAS-VARIANCE TRADEOFF

Bias

— due to wrong assumptions e.g. data is linear when it is actually quadratic.

— A high-bias model is most likely to **underfit** the training data.

Variance

— data dependency

— model's excessive sensitivity to small variations in the training data.

— A model with many degrees of freedom will **overfit** the training data.

— *Irreducible error*

— noisiness of the data

— Change, improve the measurement process

REGULARIZATION

UCL
SCHOOL OF
MANAGEMENT

REGULARIZATION

— Reduce the number of parameters

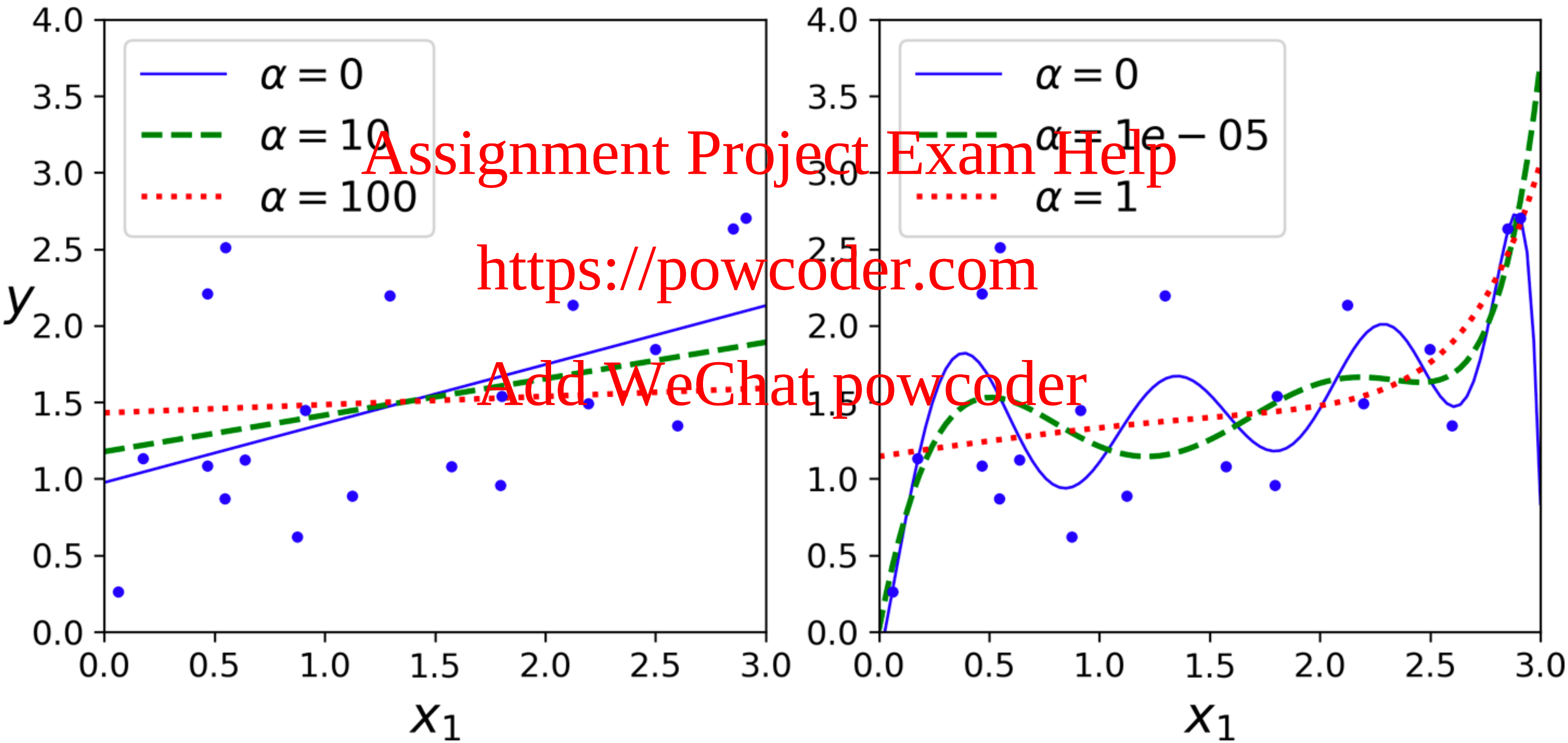
— Ridge Regression

— Lasso Regression

— Elastic Net

RIDGE (REGULARIZED) REGRESSION

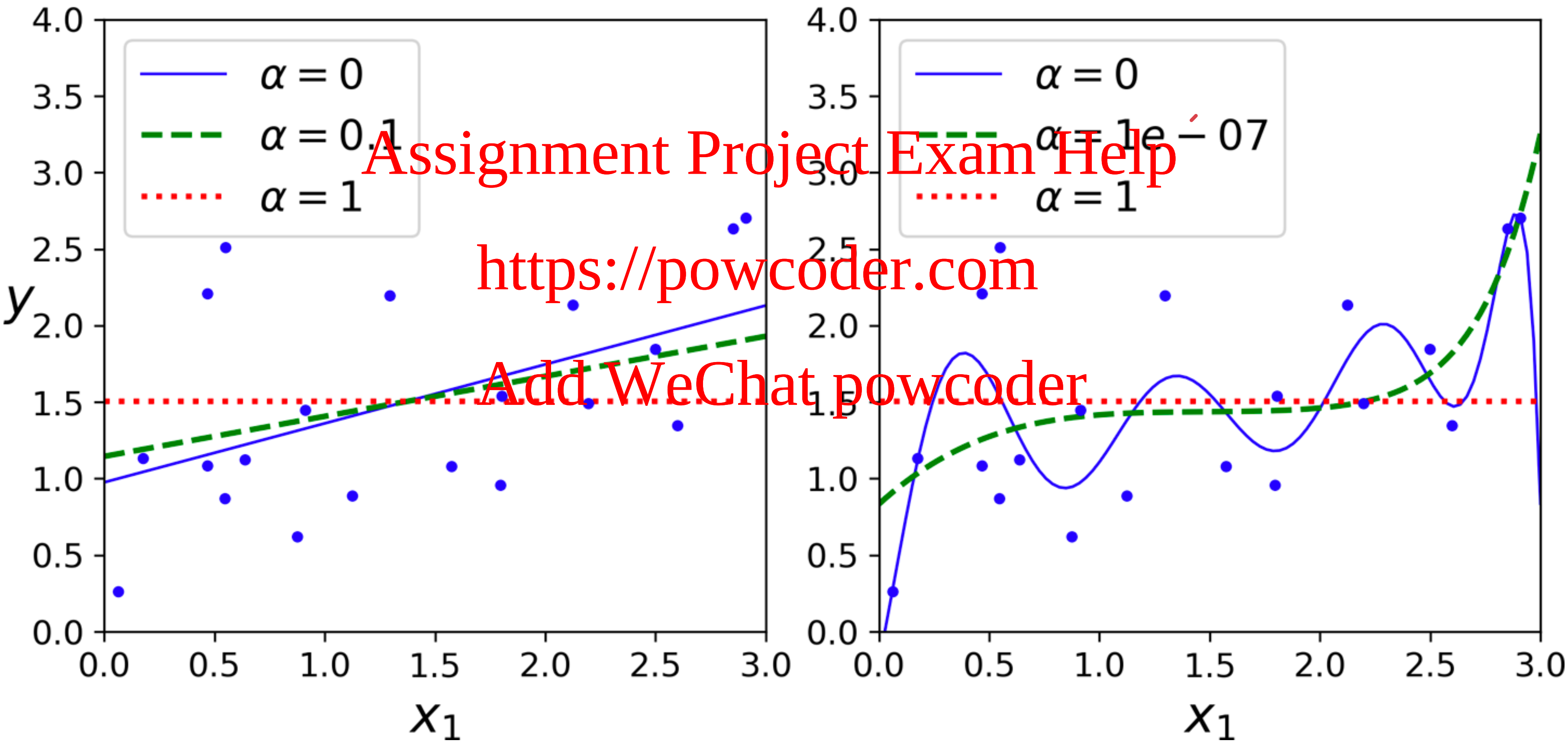
$$J(\boldsymbol{\theta}) = \mathrm{MSE}(\boldsymbol{\theta}) + \alpha \frac{1}{2} \sum_{i=1}^{n} \theta_i^{\,2}$$



LASSO (REGULARIZED) REGRESSION

UCL SCHOOL OF MANAGEMENT

$$J(\boldsymbol{\theta}) = \mathrm{MSE}(\boldsymbol{\theta}) + \alpha \sum_{i=1}^{n} |\theta_i|$$



LECTURE 2 TERM 2:

MSIN0097

UCL SCHOOL OF MANAGEMENT

PREDICTIVE ANALYTICS

VIDEO 6: REGRESSION

A P MOORE