

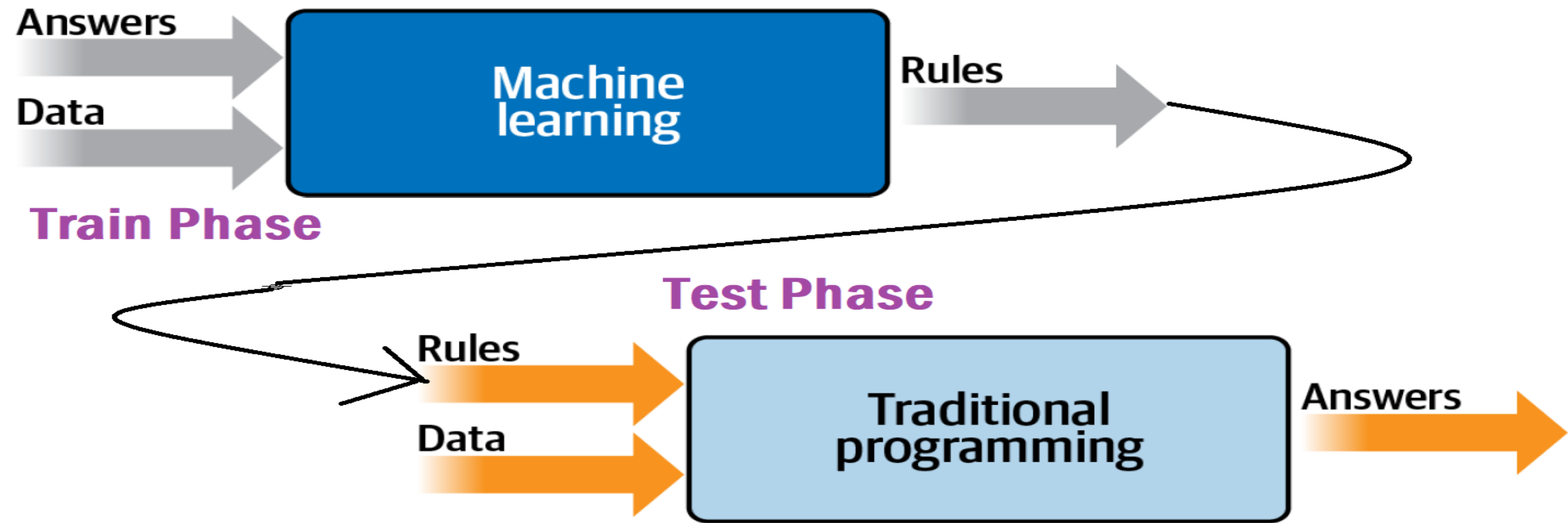


# Lecture 2 & 3: Machine Learning Perspectives

# Recap

- Compare & Contrast
  - Traditional programming with machine learning
  - How ML scales by learning the pattern
- Brief history of AI from ML perspective
- Compare & Contrast AI and ML
- Different types of Machine Learning

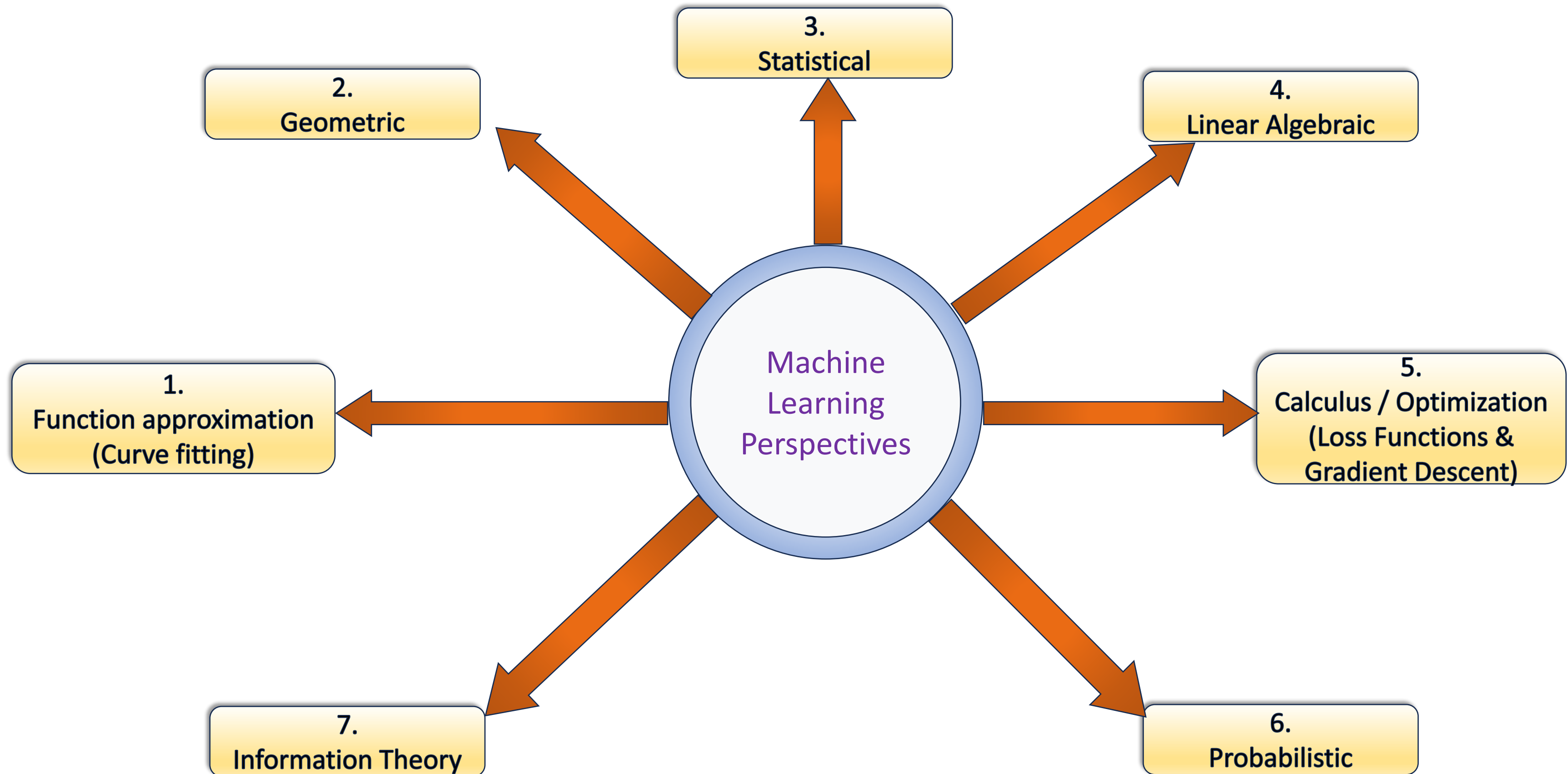
# Machine Learning is like exam



- Success is measured on
  - How one performs in test
  - Not how one practices



# ML perspectives



# Purpose

- This overview lecture is mile wide and inch deep
- To provide a holistic perspective of ML landscape
- Reflect on this lecture & revisit at semester end
- To enable reading the menu but not to cook
- Understand intuitively WHAT, WHY, WHEN. Not HOW
- Certain ML problems are better envisioned in one way rather than another
- To understand and to convince others



# 1. Function Approximation (Curve Fitting) Perspective for regression



# Analytical function

- Straight Line  $y = 2x + 3$

Geogebra link:

<https://www.geogebra.org/calculator/bbauvftd>

- What is independent & dependent variables here?
- What is the domain and co-domain?
- Mathematical Notation -  $f : \mathbb{R} \rightarrow \mathbb{R}$
- Plane  $z = 2x + y + 3$   $f : \mathbb{R}^2 \rightarrow \mathbb{R}$
- In ML,  $y$  is reserved for target variable
- Therefore we write  $y = 2x_1 + x_2 + 3$
- Line, Plane, Hyperplane – How do we generalize

Geogebra link:

<https://www.geogebra.org/calculator/nsvgpmgz>

# Going in the reverse direction is Machine learning

- Given data, deduce the function
- Boston dataset – Record as row, Feature as column

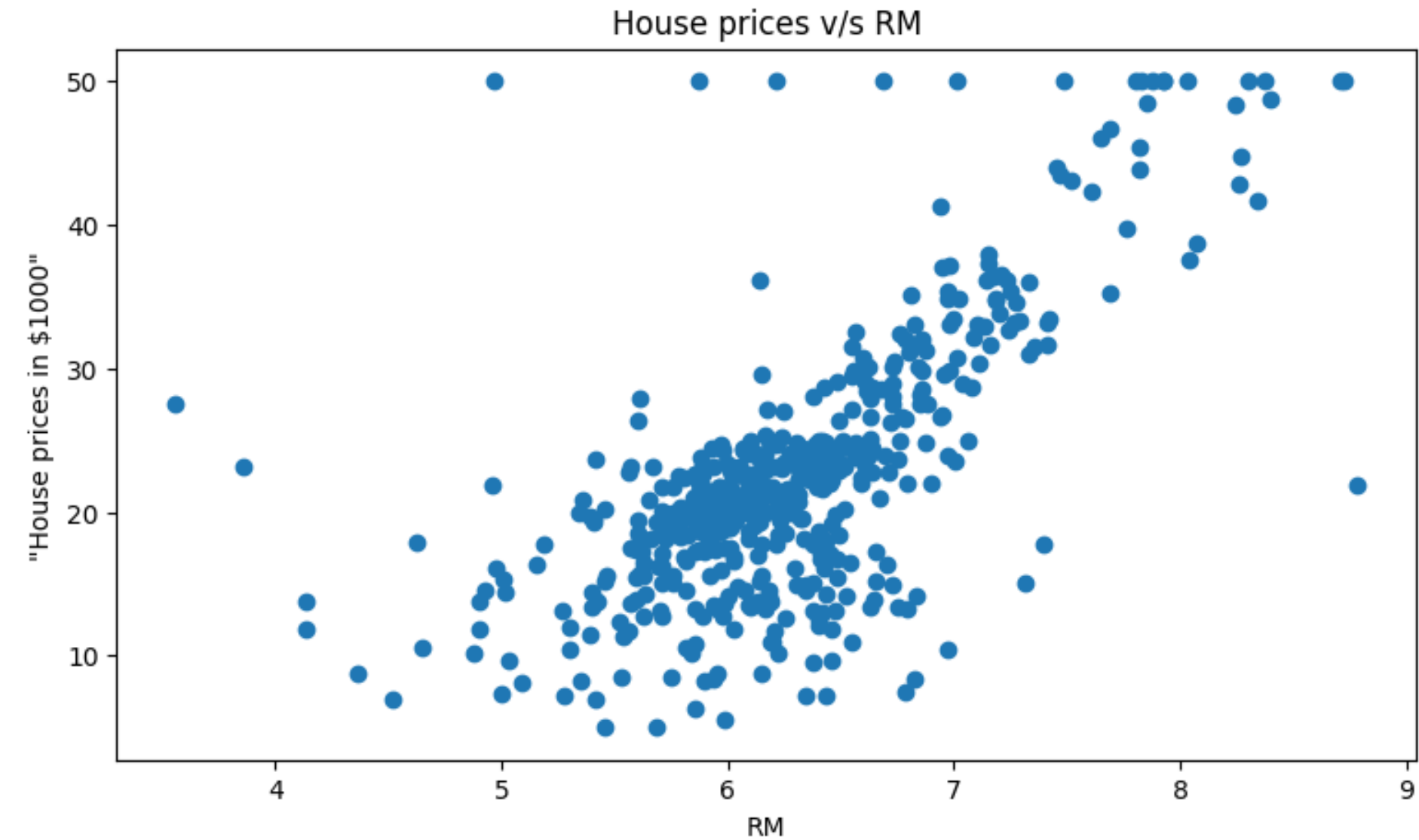
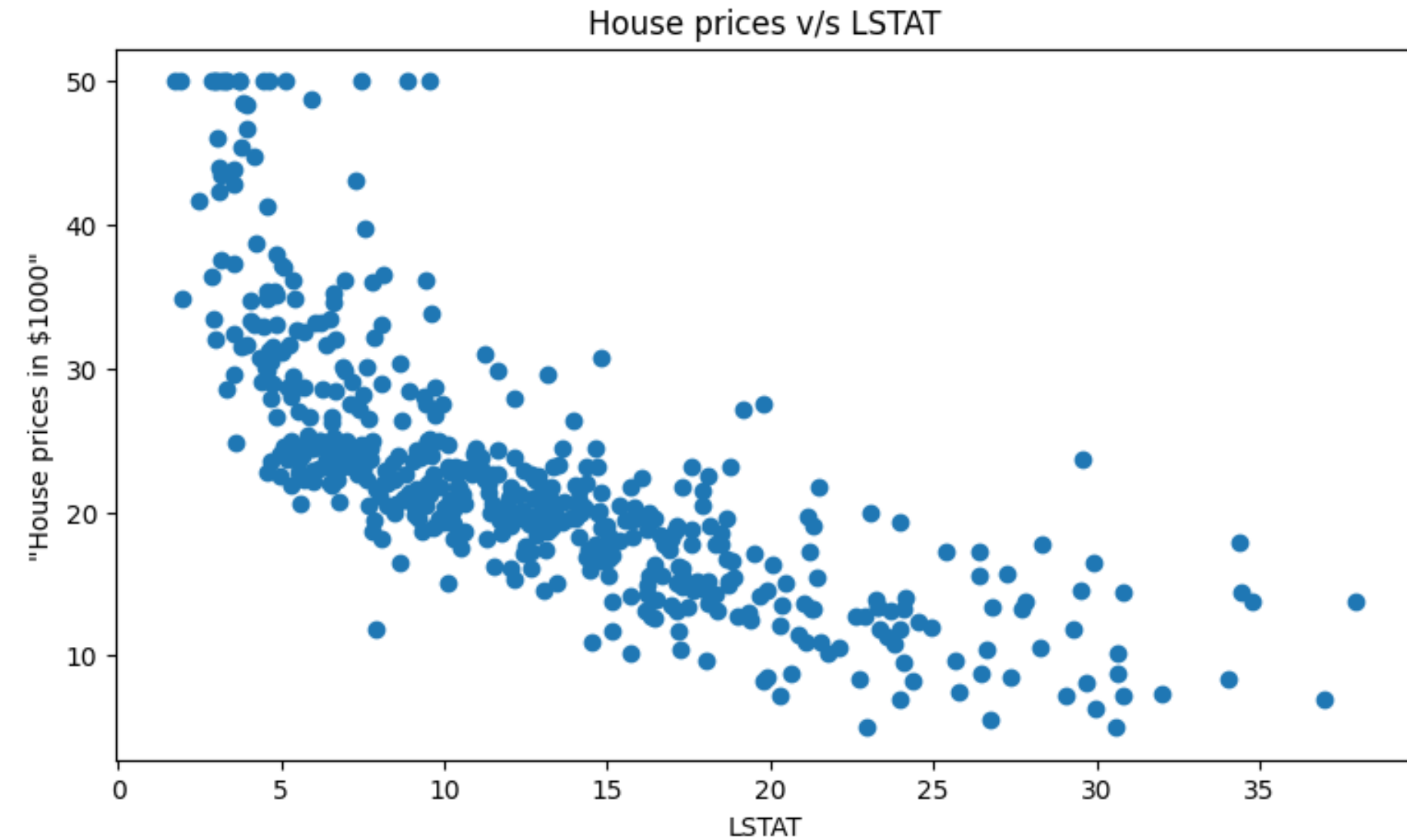
	CRIM	AGE	RM	LSTAT	PRICE
0	0.00632	65.2	6.575	4.98	24.0
1	0.02731	78.9	6.421	9.14	21.6
2	0.02729	61.1	7.185	4.03	34.7
3	0.03237	45.8	6.998	2.94	33.4
4	0.06905	54.2	7.147	5.33	36.2

- What is x and y here? (or rather X and y)
- What are alternate names for X and y?

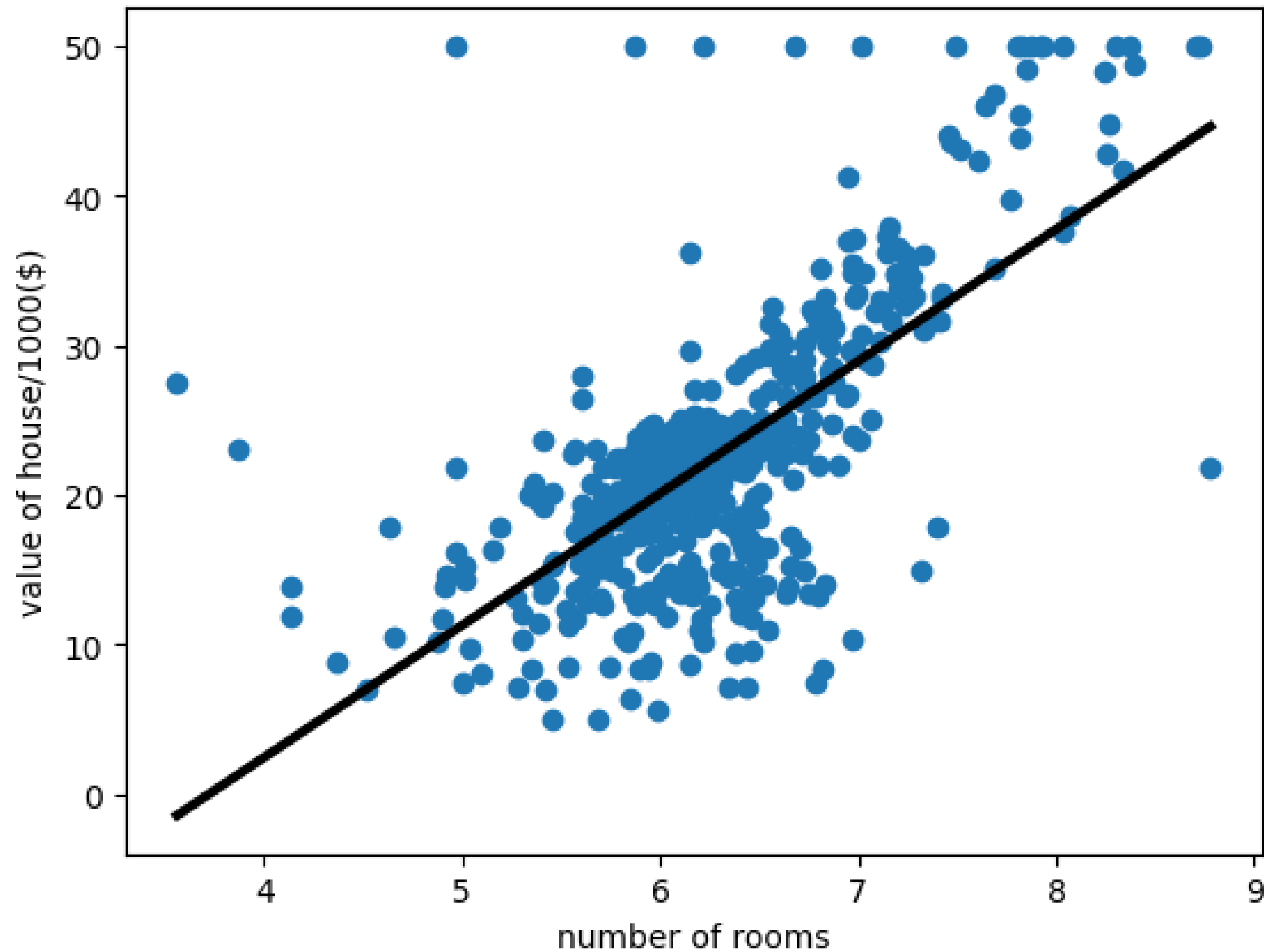


# Plotting x and y

- Let us take one variable at a time



# Fitting a line



- How is the line chosen?

# Analytical function

- 2D - Line  $y = 2x + 3$   $f : \mathbb{R} \rightarrow \mathbb{R}$
- 3D - Plane  $y = 2x_1 + x_2 + 3$   $f : \mathbb{R}^2 \rightarrow \mathbb{R}$
- $\geq 4$ D Hyperplane
- How do we generalize?



# Predicted function (hypothesis function)

$$y = 2x_1 + x_2 + 3$$

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}$$

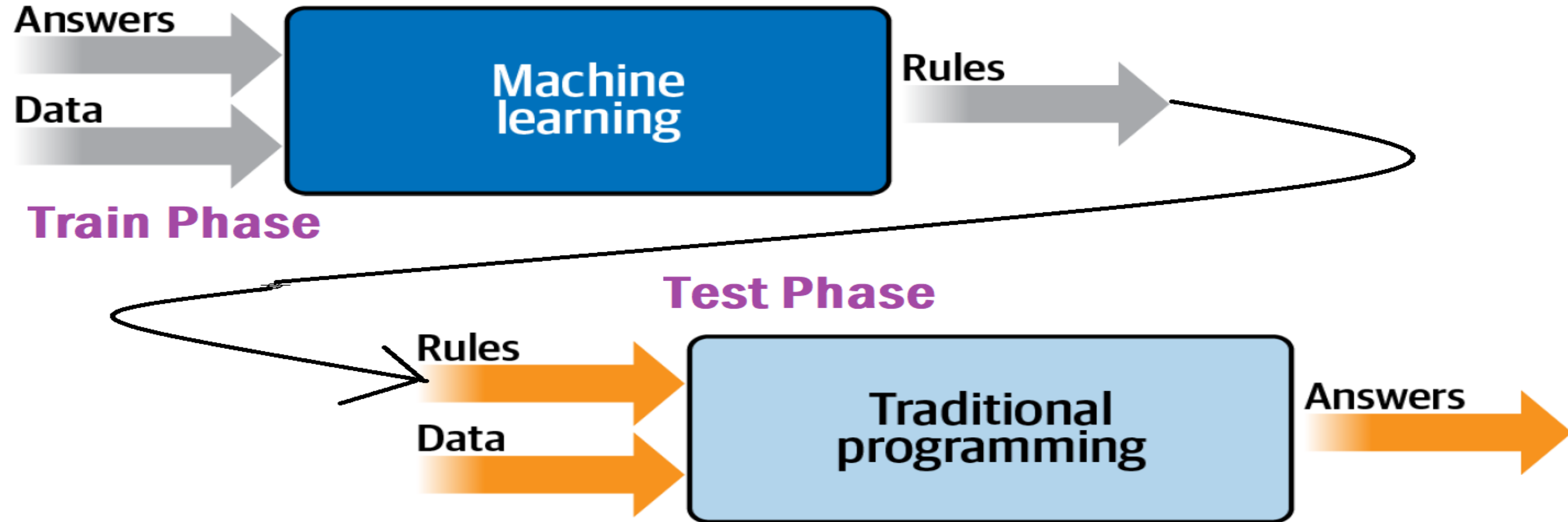
$$w = \underbrace{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}_{\text{Coefficients/Weights vector}}$$

$$x = \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\text{Features vector}}$$

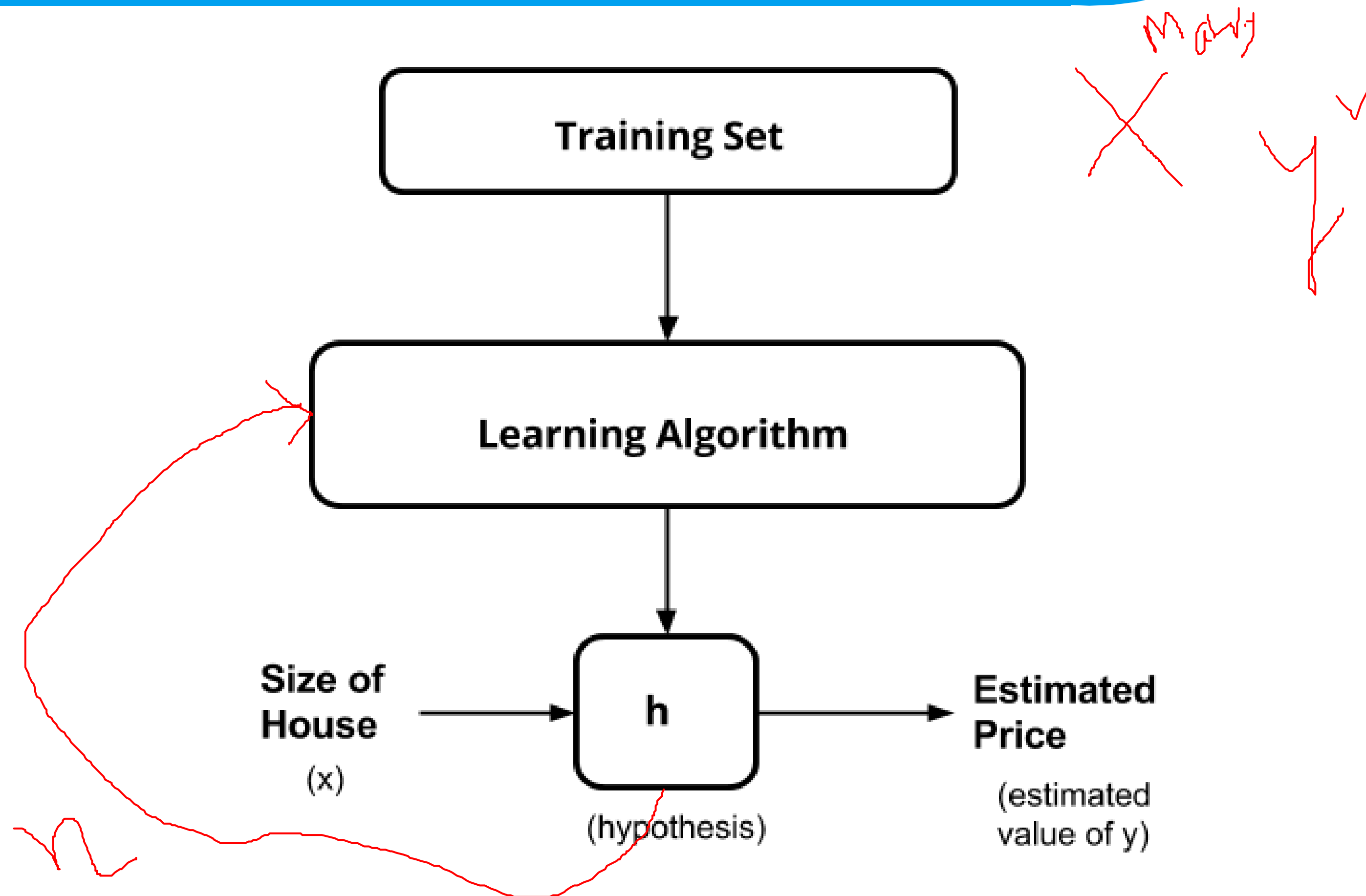
$$w^T x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$y = w^T x + b$$

# One last look

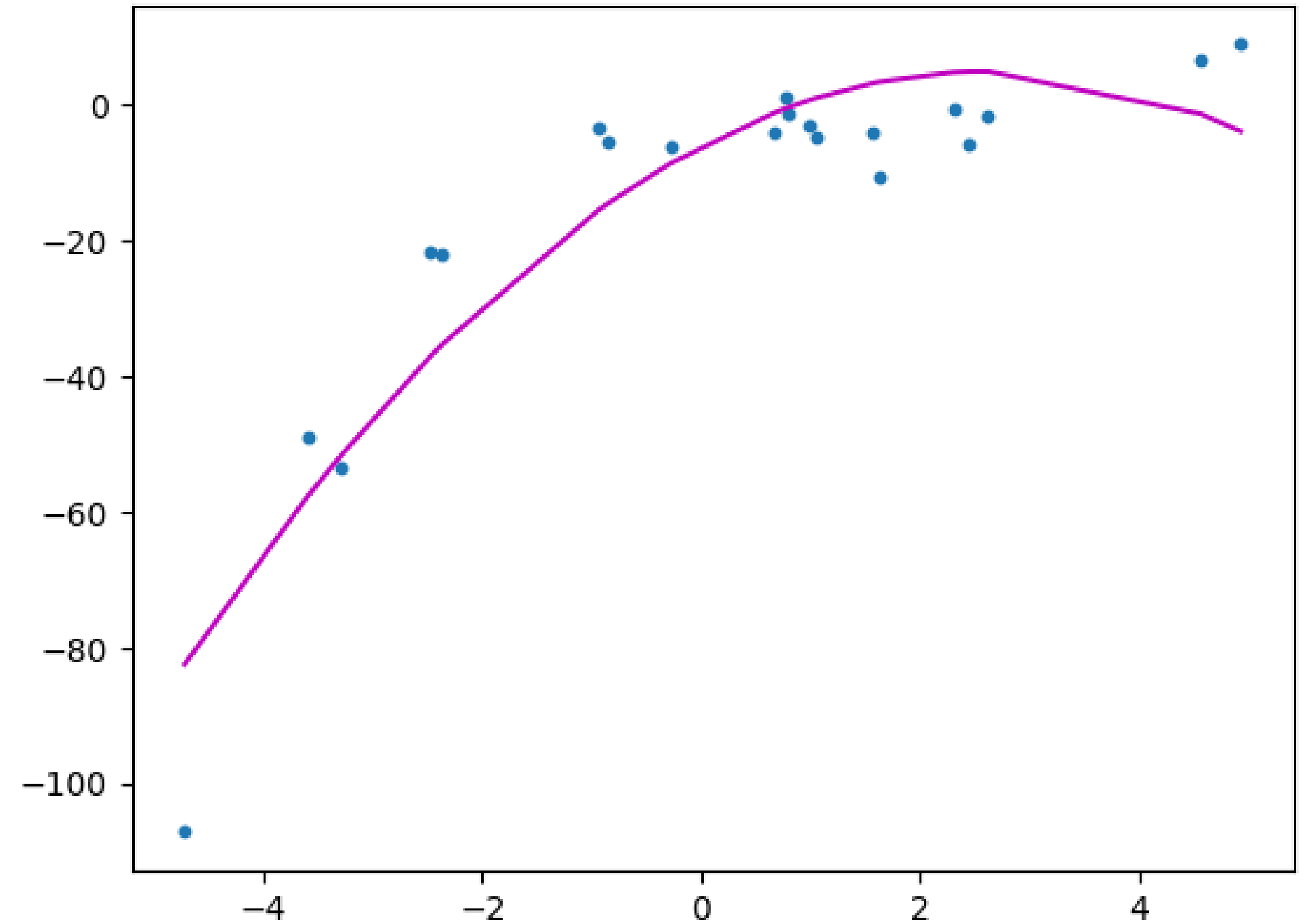
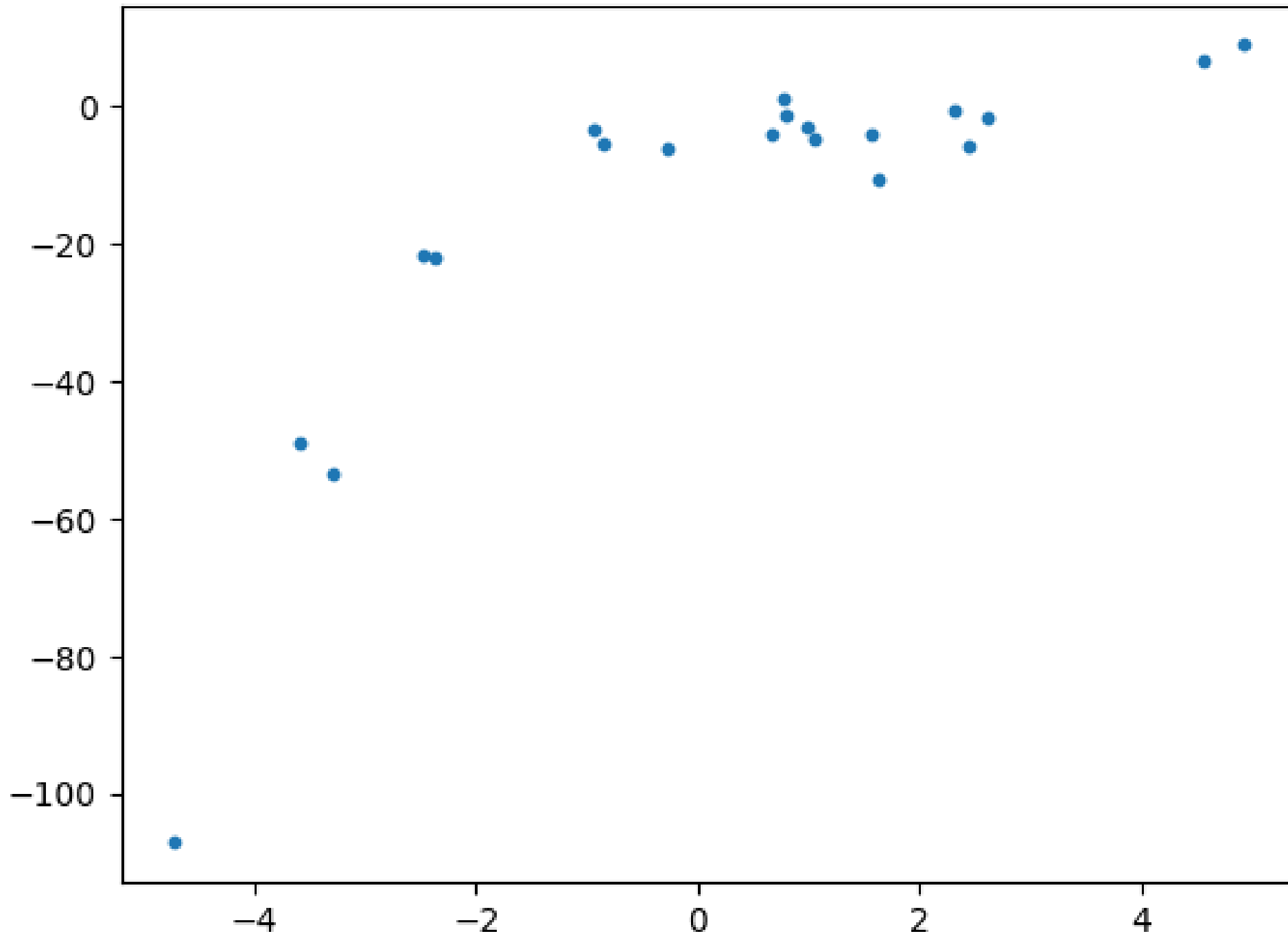


# Learning algorithm & hypothesis function

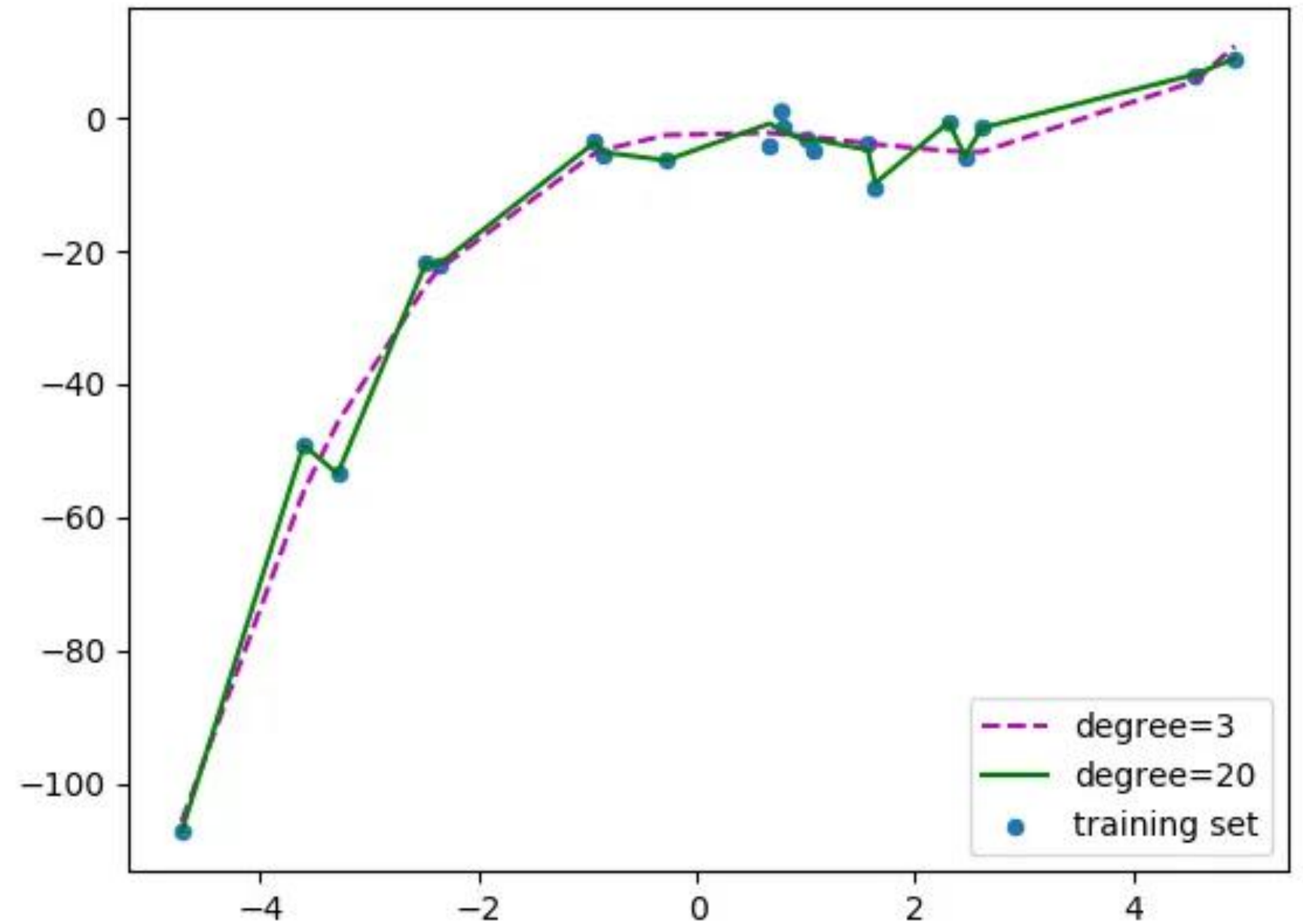
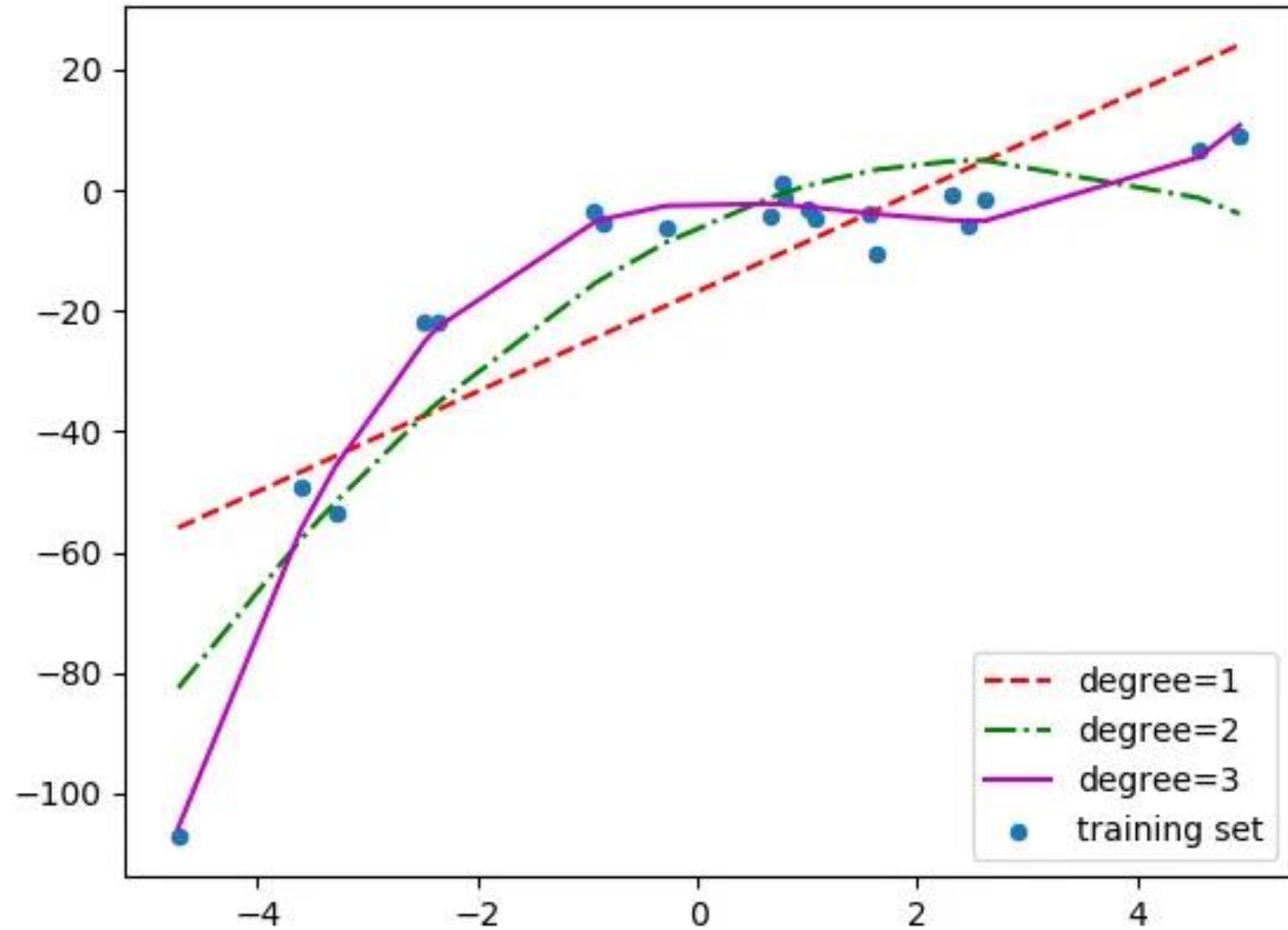




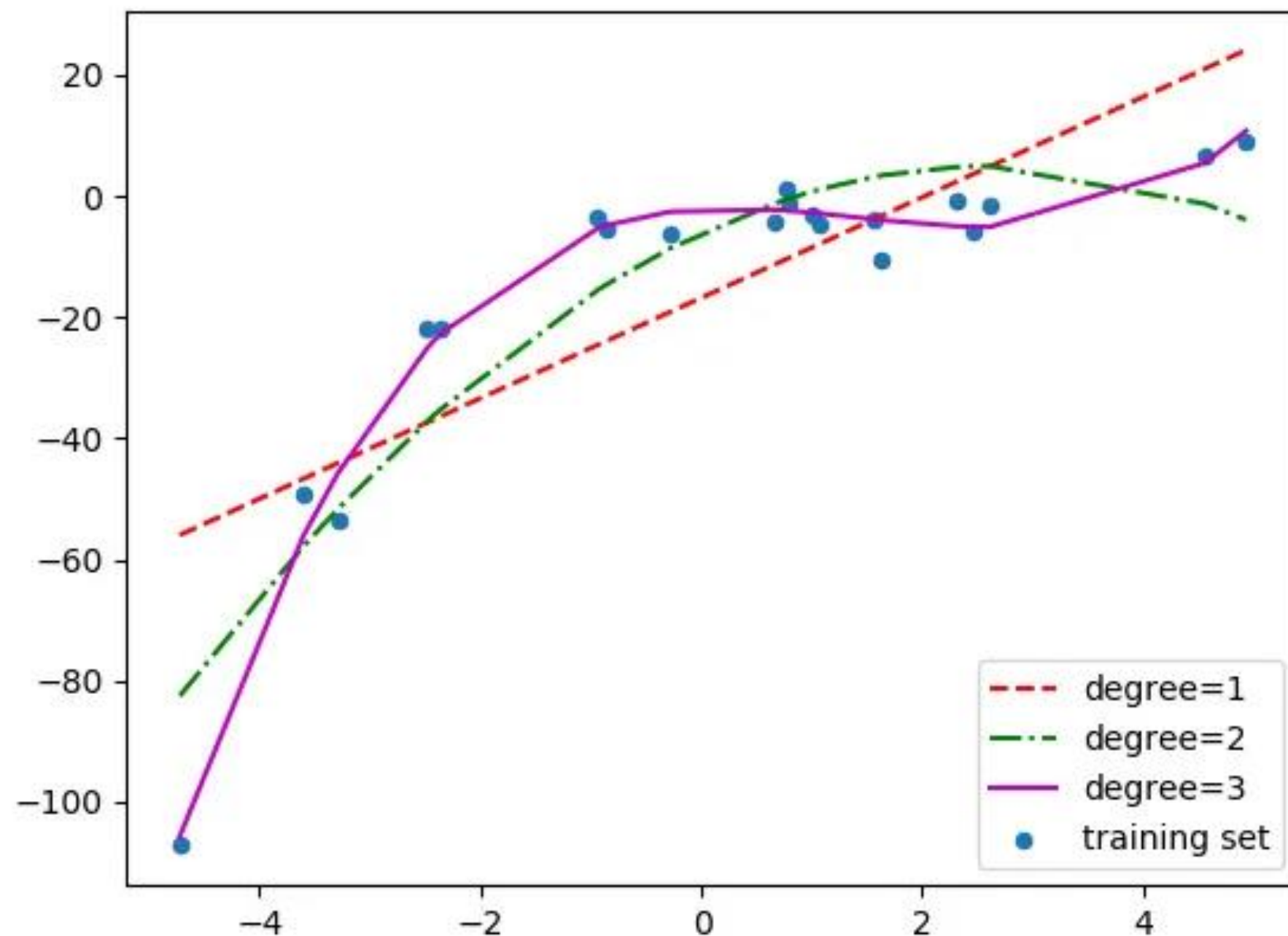
# Fitting a curve



# How many degrees is good?



# Polynomial Regression



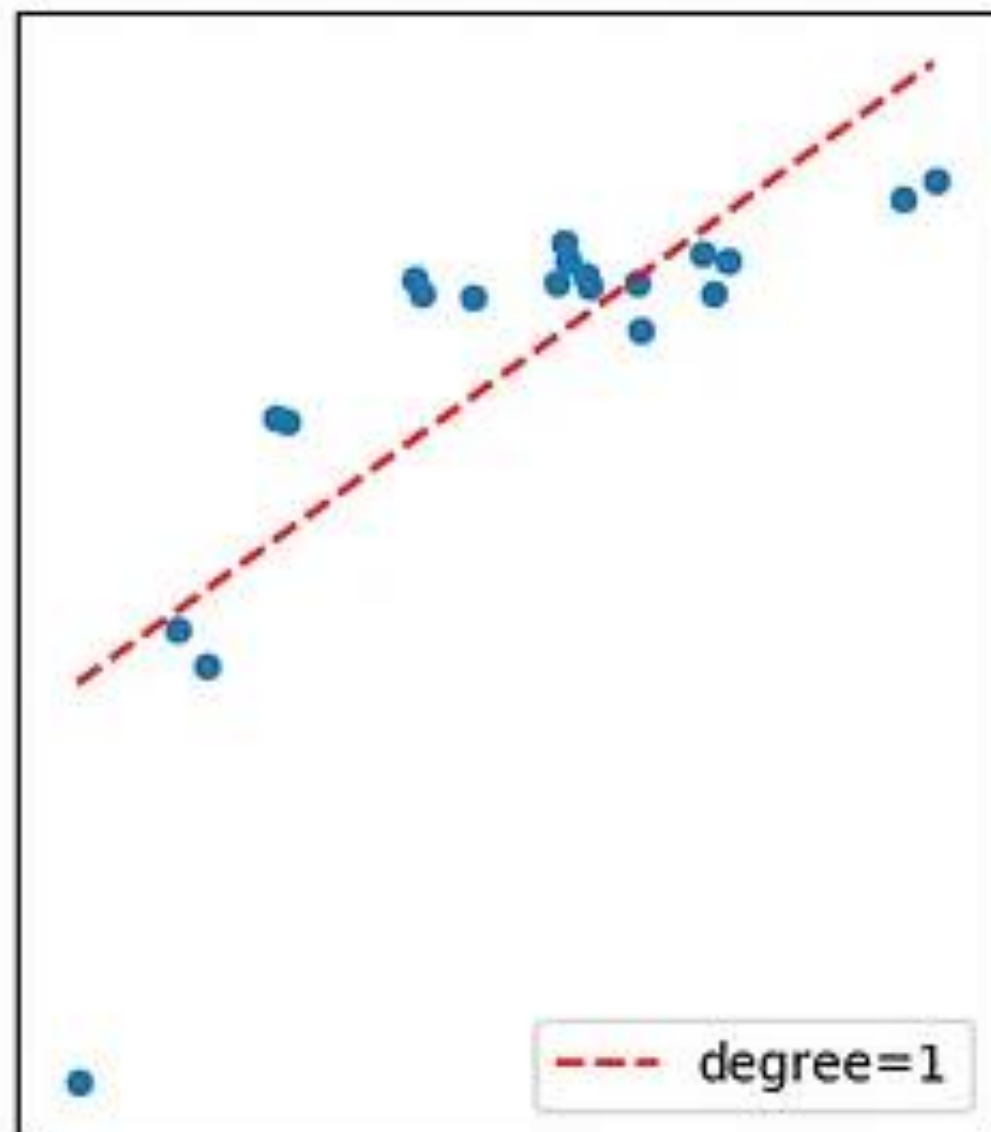
$$y = \mathbf{w}^T \mathbf{x} + b$$

$$y = w_1 x + w_2 x^2 + w_3 x^3 + b$$

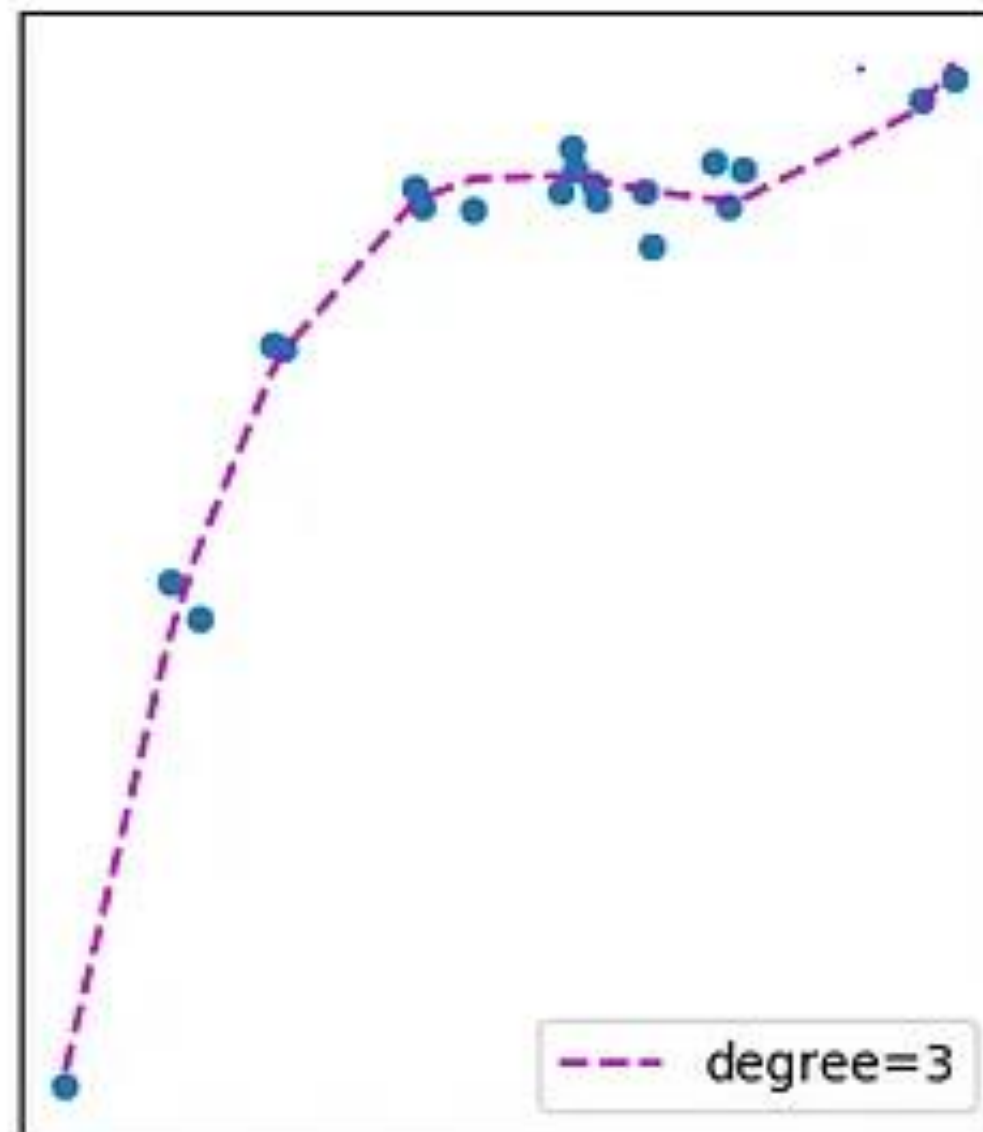
$$y = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}^T \underbrace{\begin{bmatrix} x \\ x^2 \\ x^3 \end{bmatrix}}_{\text{Features}} + b$$



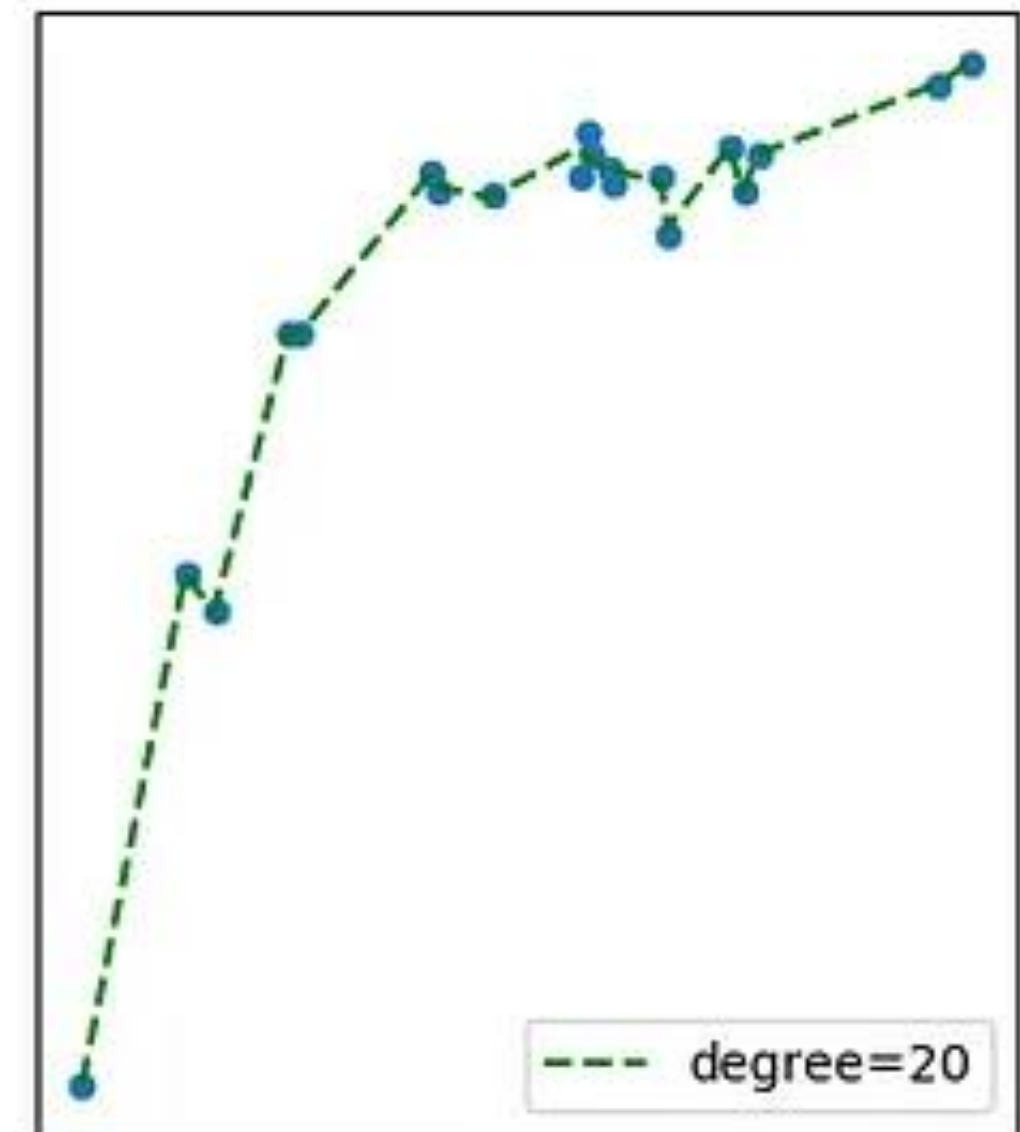
# Underfitting & overfitting



Underfit  
High Bias  
Low Variance



Correct Fit  
Low Bias  
Low Variance



Overfit  
Low Bias  
High Variance

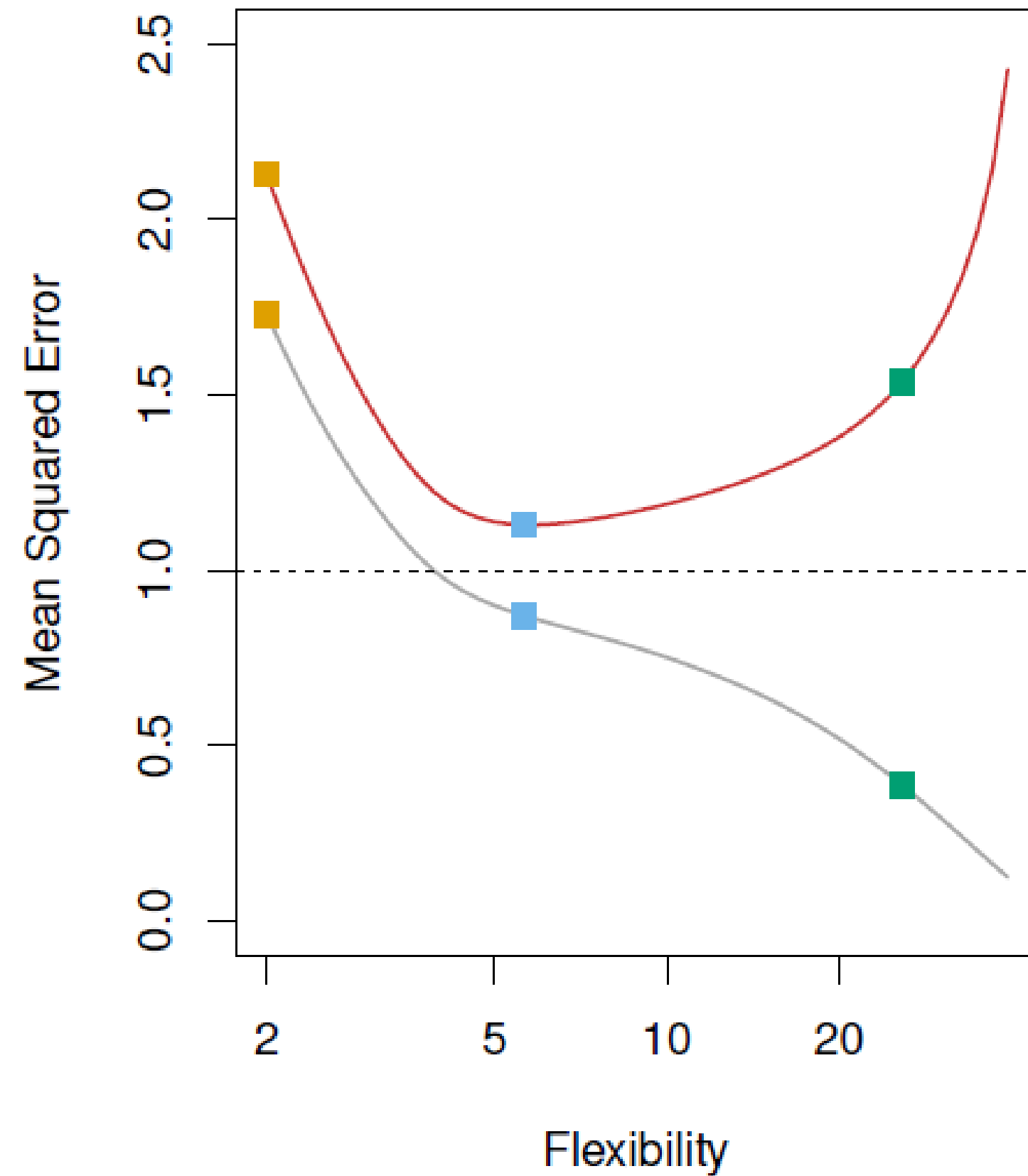
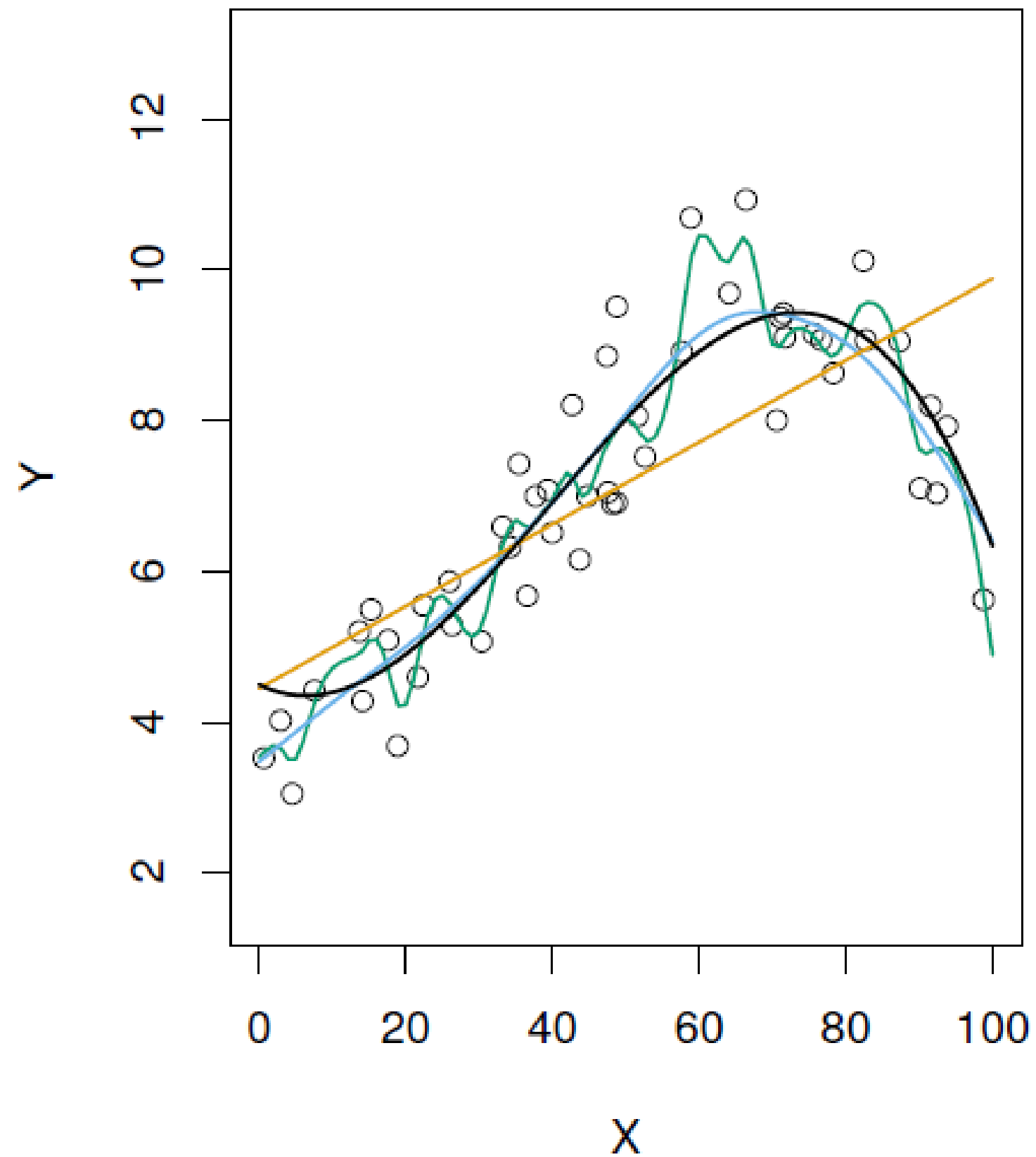
# Learning pattern v/s noise

- Overfitting likely tends to learn from noise
- Happiness Index

Country	Per Capita Income	Happiness Index
Finland	25.0	7.82
Denmark	25.0	7.64
Iceland	24.0	7.56
Switzerland	27.0	7.51
Netherlands	22.0	7.42
Luxembourg	22.0	7.40
Sweden	21.0	7.38
Norway	21.0	7.37
Israel	20.0	7.36
Newzealand	20.0	7.20

Country	Per Capita Income	Happiness Index
Somaliland	5.1	3.75
Tanzania	5.0	3.70
Zimbabwe	4.0	3.00

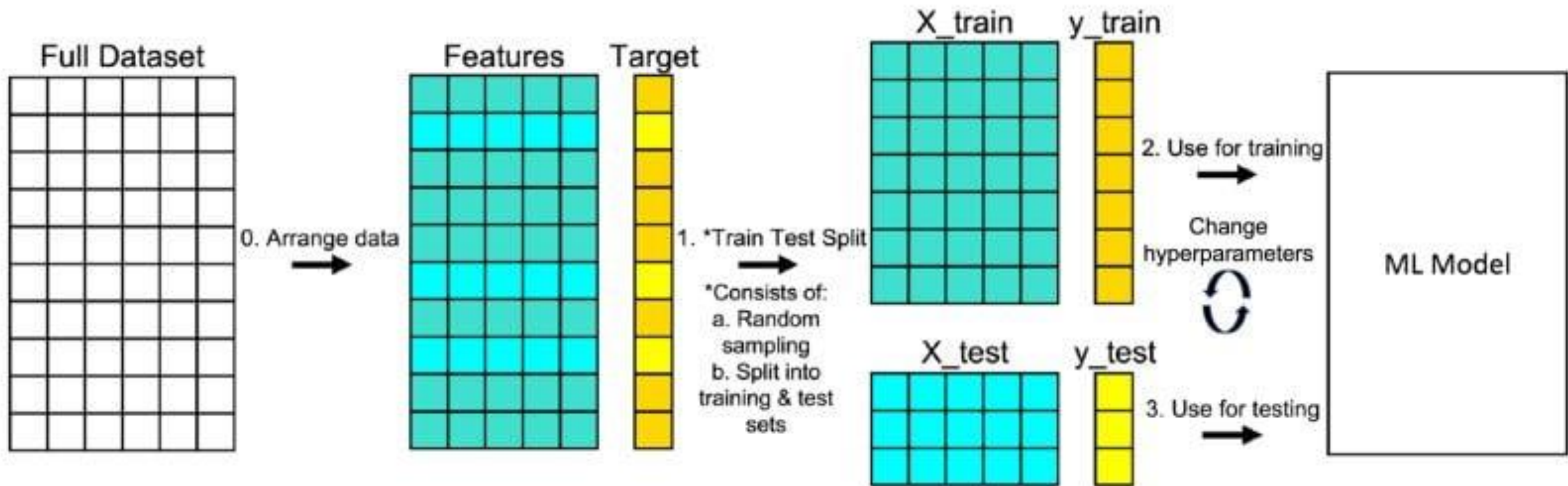
# Bias Variance Trade-off





# Train Test Split

- Recall: ML is like exam performance



# Recap

	CRIM	AGE	RM	LSTAT	PRICE
0	0.00632	65.2	6.575	4.98	24.0
1	0.02731	78.9	6.421	9.14	21.6
2	0.02729	61.1	7.185	4.03	34.7
3	0.03237	45.8	6.998	2.94	33.4
4	0.06905	54.2	7.147	5.33	36.2

- How many dimensional data here?
- Is this a regression or classification problem?

# Takeaways

- Curve fitting is logical way of looking at regression problems
- Line, plane & hyperplane - one form  $y = w^T x + b$
- Learning algorithm & hypothesis function
- How many degrees is just right
- Underfitting, Overfitting and just right fit
- Bias Variance Tradeoff
- Overfitting discovered by Cross validation and error increasing beyond a point
- Overfitting learns noise





QUESTIONS





# Thank You!