week 1

السنة الخامسة – هندسة المعلوماتية / الذكاء الصنعي

مقرر التعلم التلقائي

مقدمة إلى التعلم التلقائي (تعلم الآلة)
Introduction to Machine Learning

د. رياض سنبل

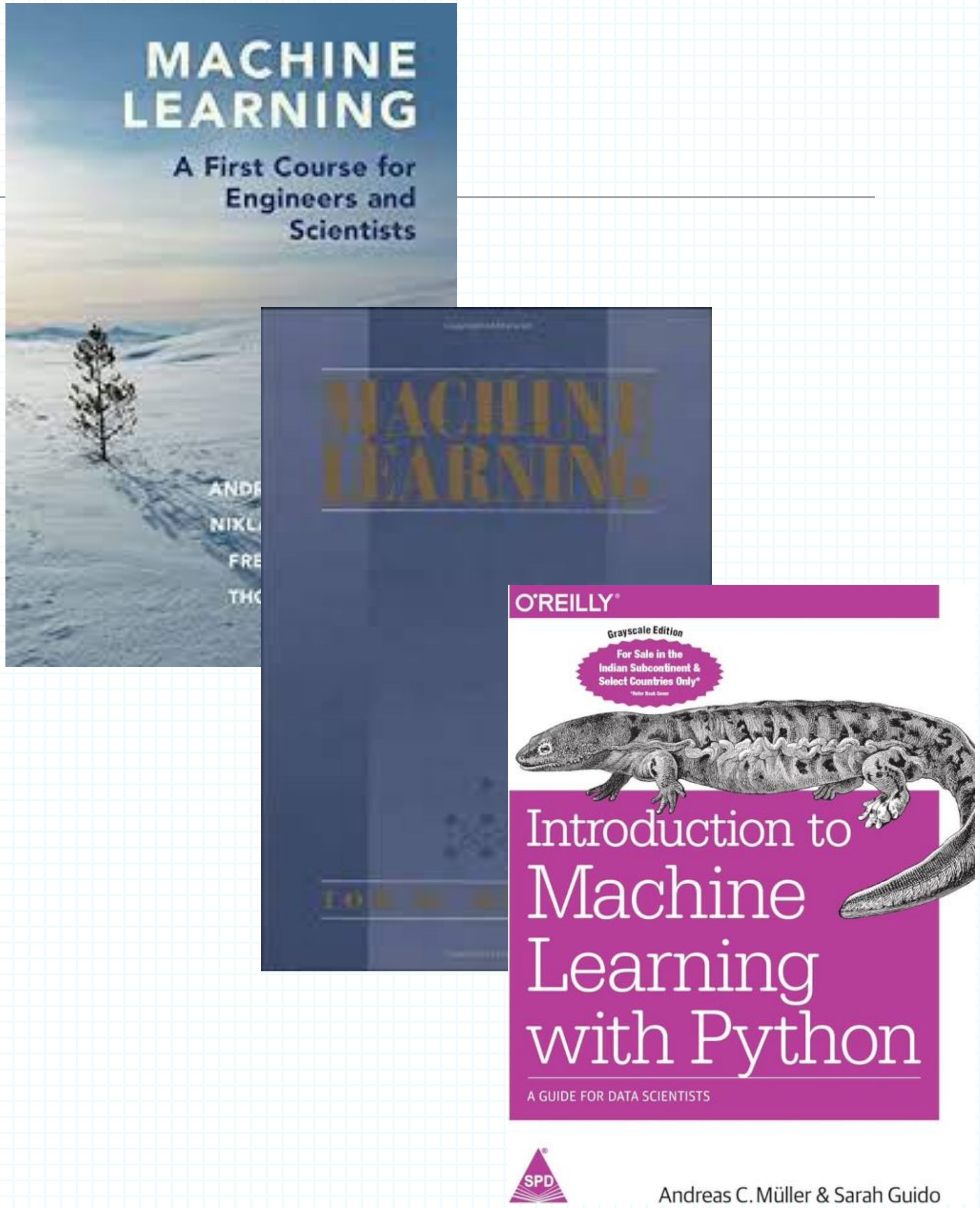
Access Course Materials

Road Map: *From Classical ML to Cutting-Edge AI*

1) Introduction to Machine Learning: Basic Concepts
2) Decision Trees
3) Estimation Strategy and Evaluation Metrics
4) Feature Engineering
5) Support Vector Machines (SVM)
6) KNN, Naive Bayes, etc
7) Ensemble Methods
8) Introduction to Deep Learning
9) Optimization in Deep Neural Networks
10) Generalization and Regularization in Deep Neural Networks
11) CNN, RNN, LSTM
12) Transformers
13) LLMs

Outline of the course

- **Textbooks:**
- *Lindholm, A., Wahlström, N., Lindsten, F. and Schön, T.B., 2022. Machine learning: a first course for engineers and scientists. Cambridge University Press.*
- *Mitchell, T.M., 1997. Machine learning. McGraw Hill.*
- *Müller, A.C. and Guido, S., 2016. Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.".*

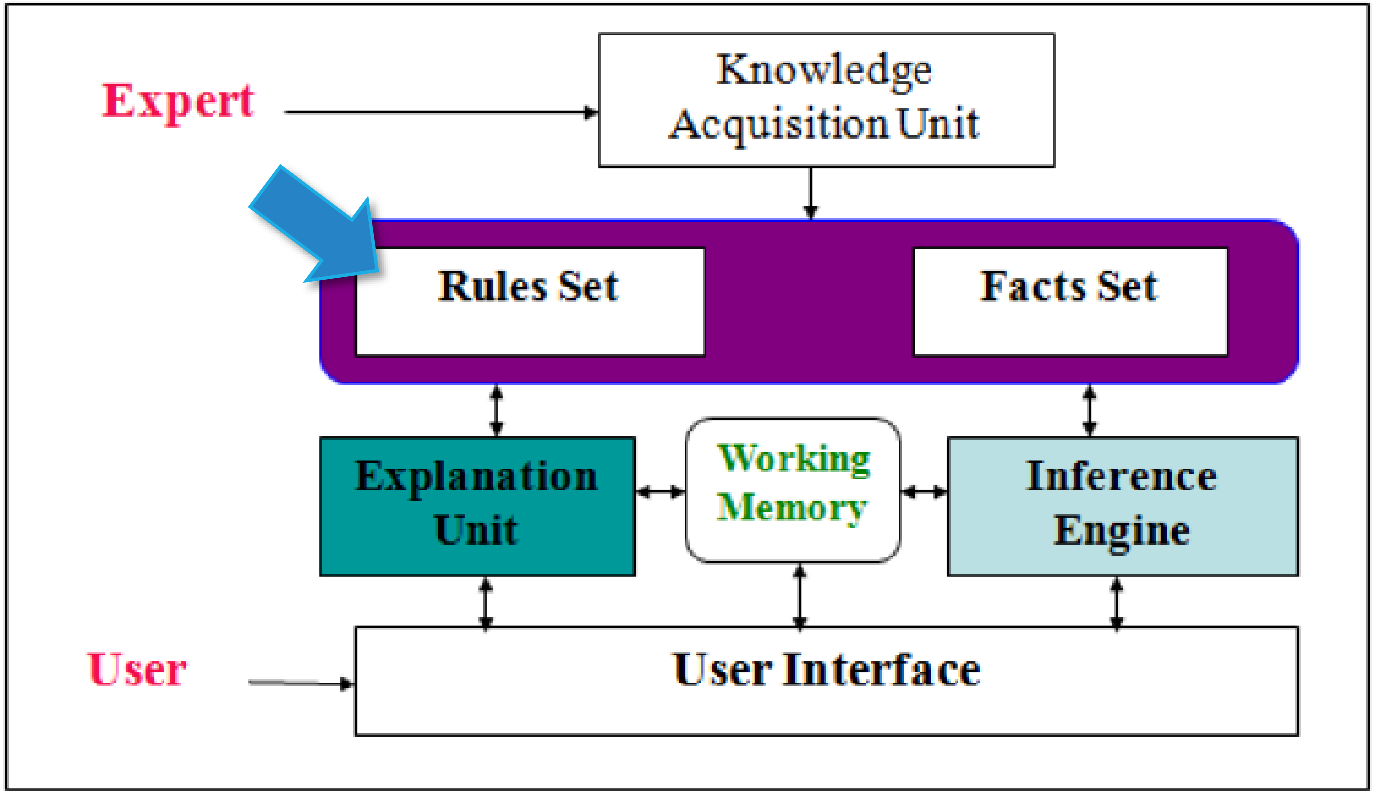


Evolution of problem-solving paradigms
Why we need Machine Learning?

- Traditional Algorithms
- Heuristics, A* algorithms, approximate algorithms
- Expert Systems
- Machine Learning

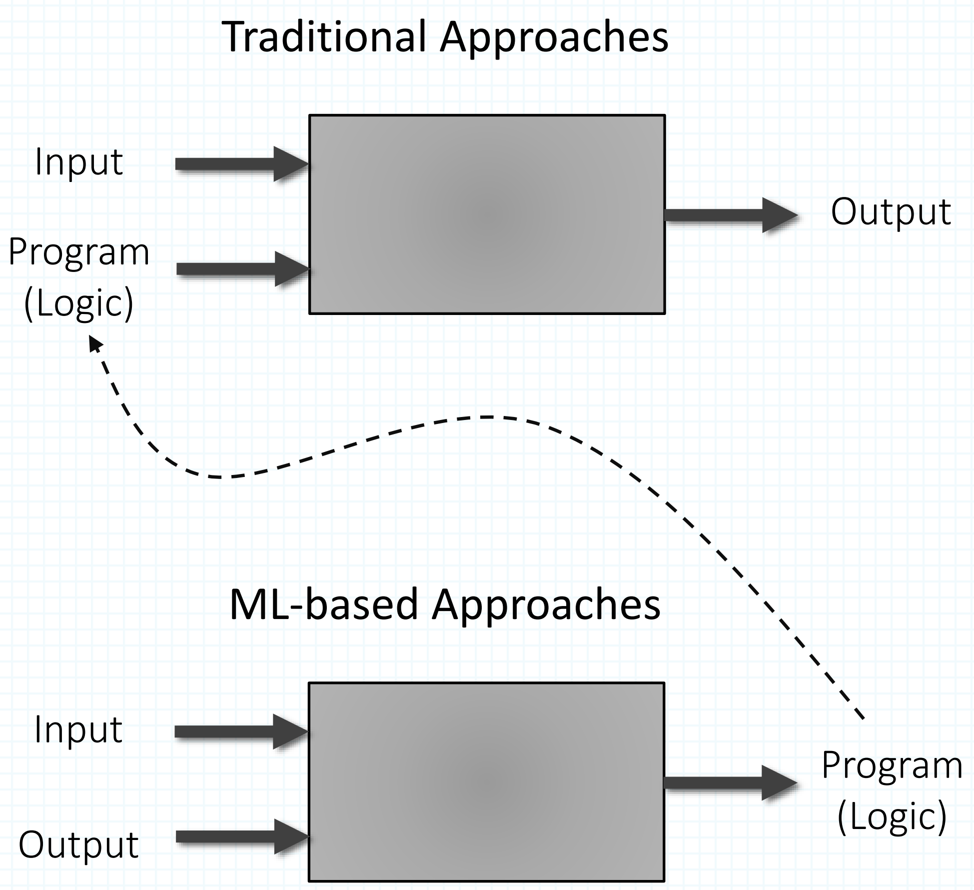
Expert Systems: A quick revision

- An expert system generally consists of four components:
  - Knowledge base (Rules)
  - Search or inference system,
  - Knowledge acquisition system,
  - User interface or communication system.



What is Machine Learning?

- Machine Learning is a type of Artificial Intelligence that provides computers with the ability to learn
- Getting computers to program themselves.



Machine Learning ≈ Looking for a Function

- At its core, **Machine Learning is about finding a function** that maps inputs to outputs:
  - Where x is your input (features, observations)
  - y is the target (label, prediction)
  - And f is the function learned from data
- Example:

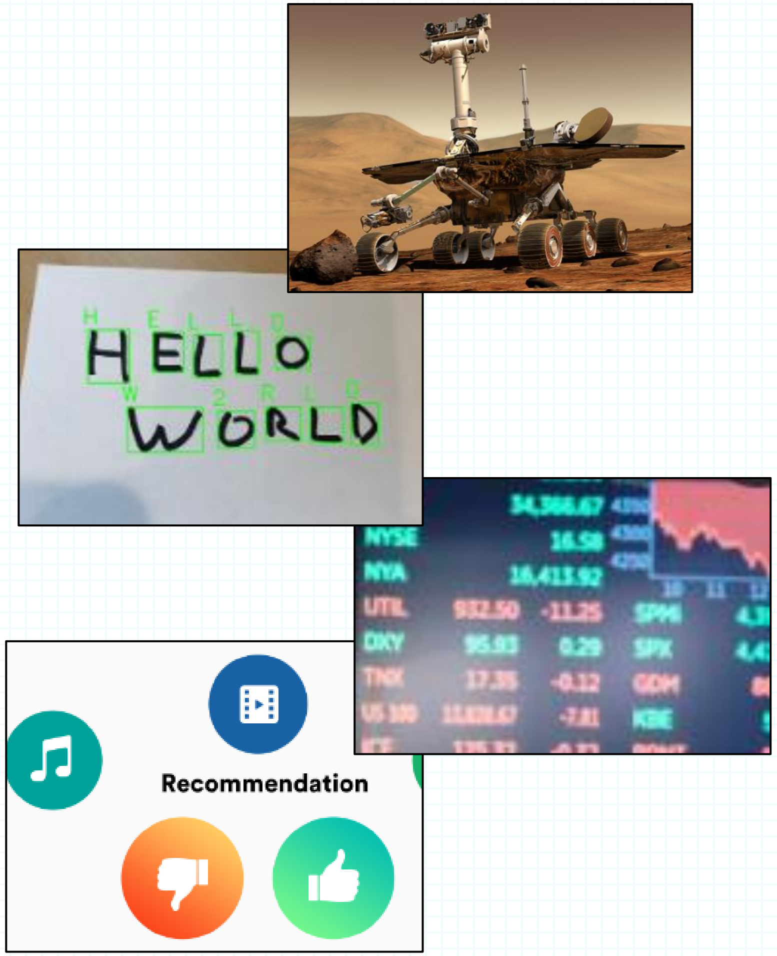
$f($  $) =$ "How are you"

$f($  $) =$ "Cat"

$f($  $) =$ "5-5" (next move)

When Do We Use Machine Learning?

- **ML is used when:**
  - Human **expertise does not exist** (navigating on Mars),
  - Humans are **unable to explain** their expertise (speech recognition, OCR)
  - Solution **changes in time** (routing on a computer network, stock market)
  - Solution needs to be **adapted to particular cases** (recommendation systems)



- **Learning is not always useful:**
  - There is no need to "learn" to calculate payroll.

Common ML Applications

- Recognizing patterns:
  - Facial identities or facial expressions.
  - Handwritten or spoken words.
  - Medical images.
  - Sentiment Analysis.
- Generating patterns:
  - Generating images or motion sequences.
  - Articles generation.
- Recognizing anomalies:
  - Unusual credit card transactions
- Prediction:
  - Future stock prices or currency exchange rates

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging

Types of Learning

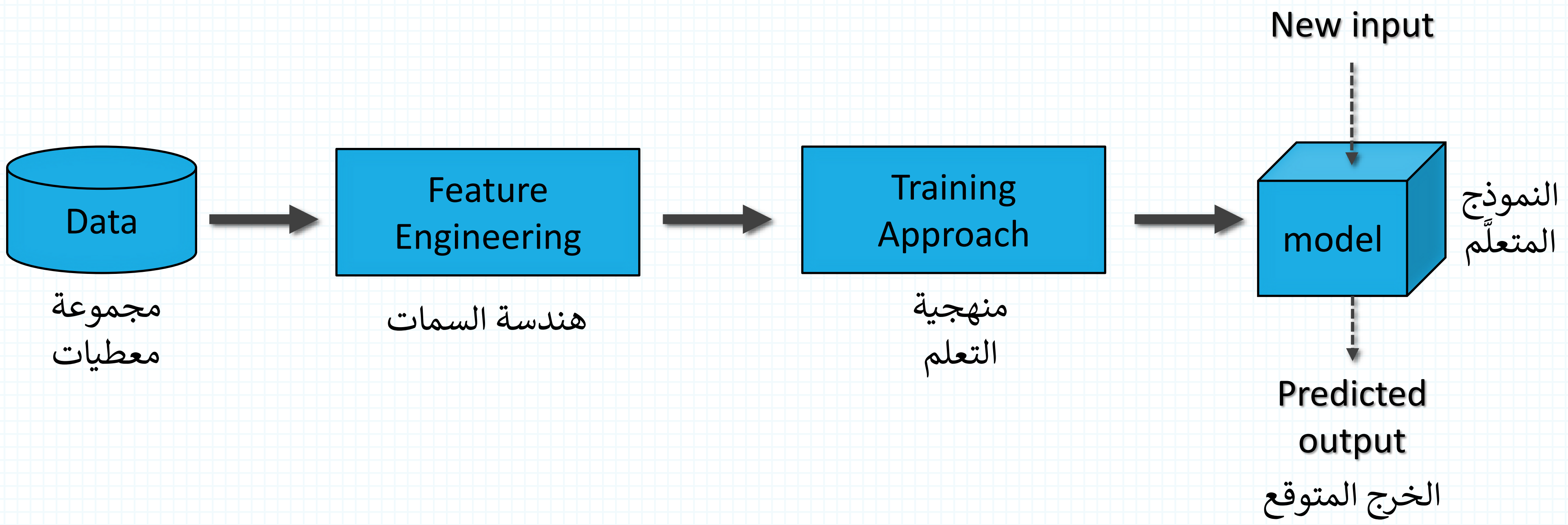
- **Supervised (inductive) learning:** Training data includes desired outputs
  - Regression: predict numerical values
  - Classification: predict categorical values, i.e., labels
- **Unsupervised learning:** Training data does not include desired outputs
  - Clustering: group data according to "distance"
  - Association: find frequent co-occurrences
- **Semi-supervised learning**
  - Training data includes a few desired outputs. It combines a small set of labeled data with a large amount of unlabeled data to improve model performance and accuracy.
- **Reinforcement learning**
  - Learn to act based on feedback/reward.
- **Self-Supervised Learning.**
- etc

Supervised Learning Techniques

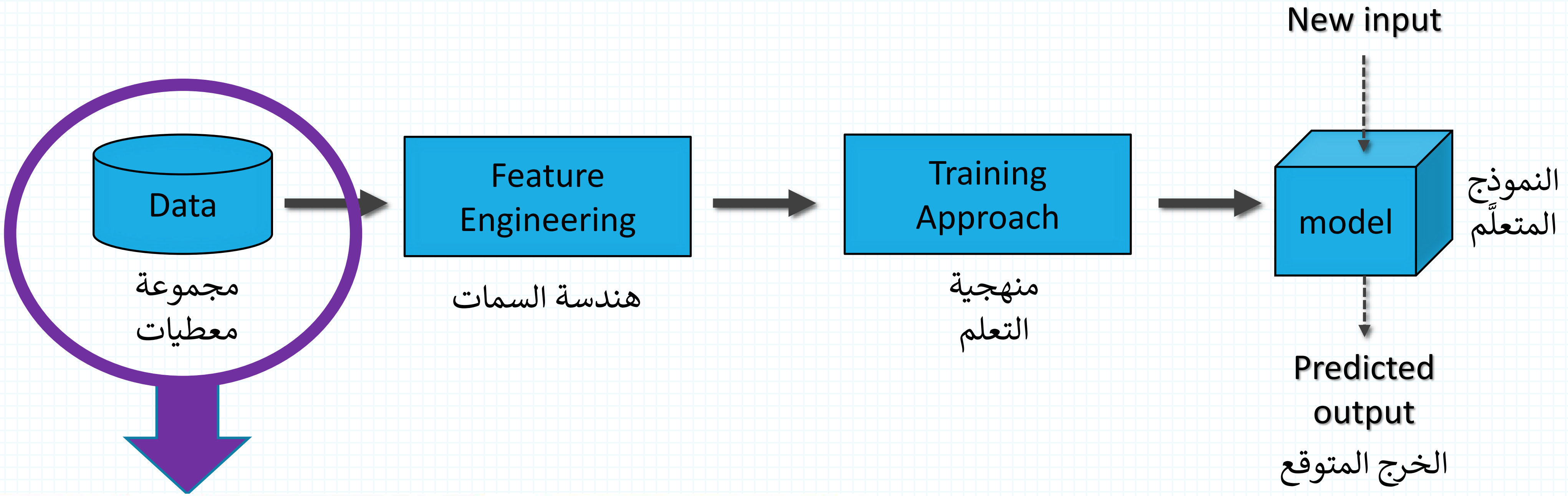
- **Numerical classifier functions**
  - Linear classifier, perceptron, logistic regression, support vector machines (SVM), neural networks
- **Parametric (probabilistic) functions**
  - Naïve Bayes, Gaussian discriminant analysis (GDA), hidden Markov models (HMM), probabilistic graphical models
- **Non-parametric (instance-based) functions**
  - *k*-nearest neighbors, kernel regression, kernel density estimation, local regression
- **Symbolic functions**
  - Decision trees, classification and regression trees (CART)
- **Aggregation (ensemble) learning**
  - Bagging, boosting (Adaboost), random forest

ML Pipeline

Traditional ML Pipeline



Traditional ML Pipeline



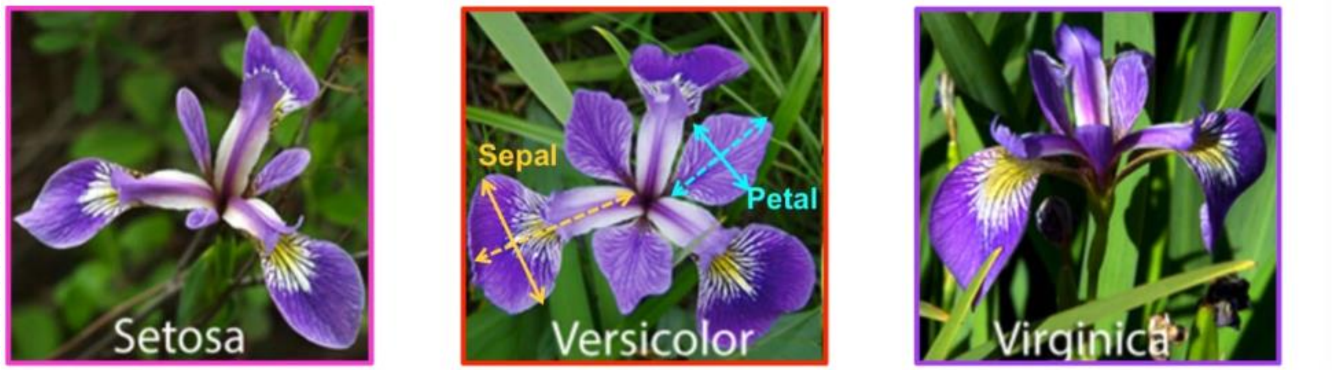
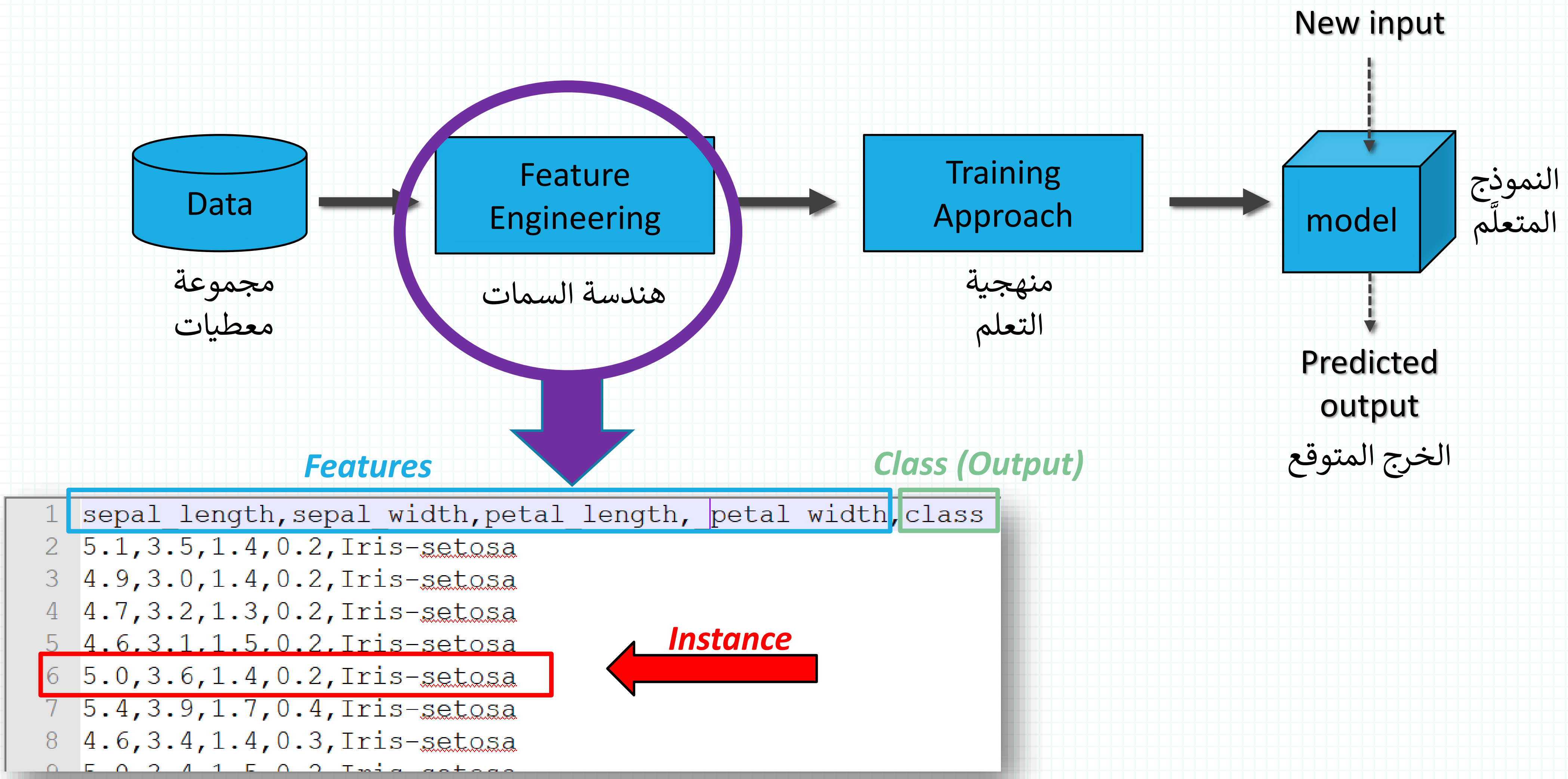
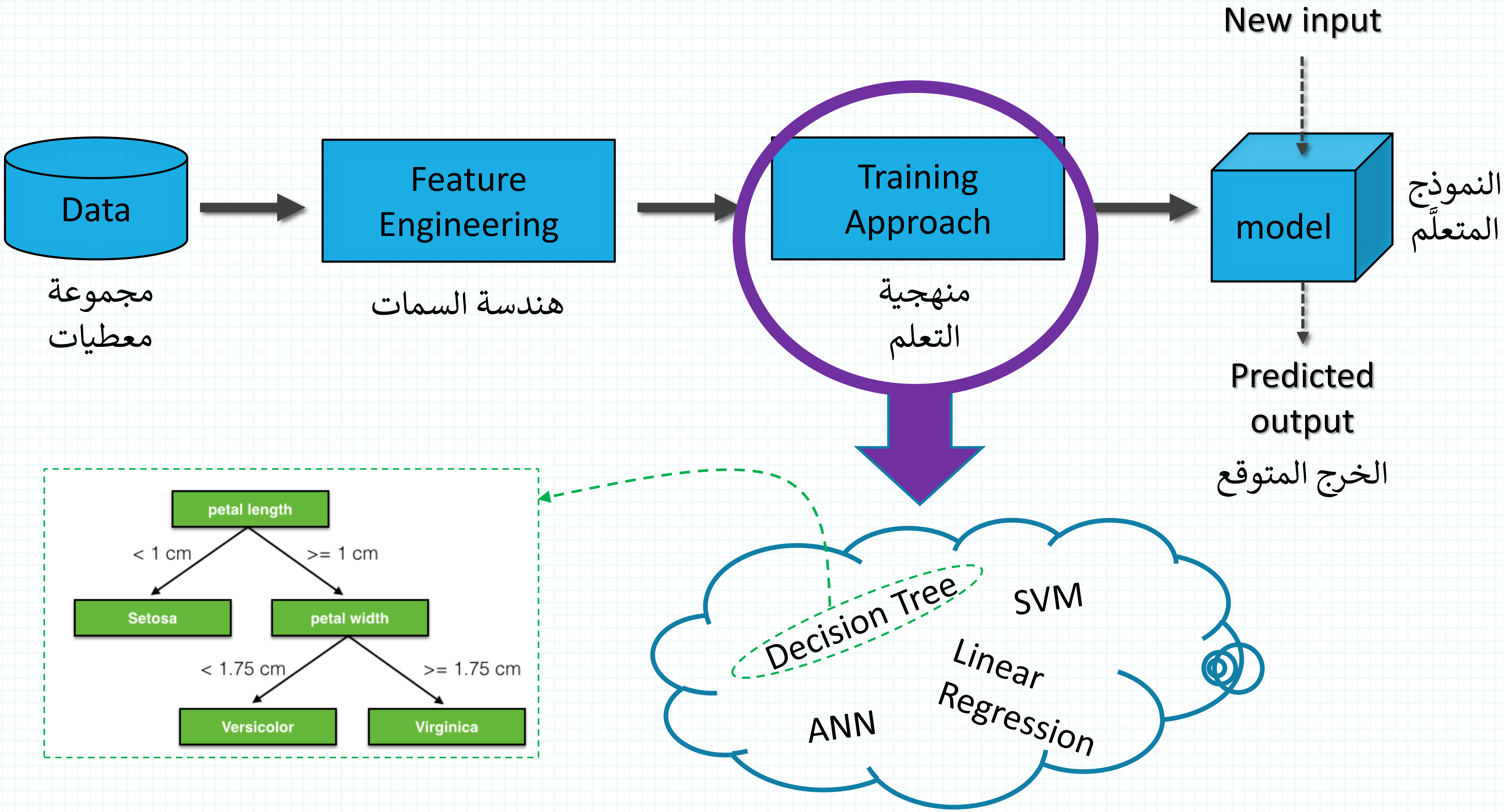


Image Source: http://suruchifialoke.com/2016-10-13-machine-learning-tutorial-iris-classification/

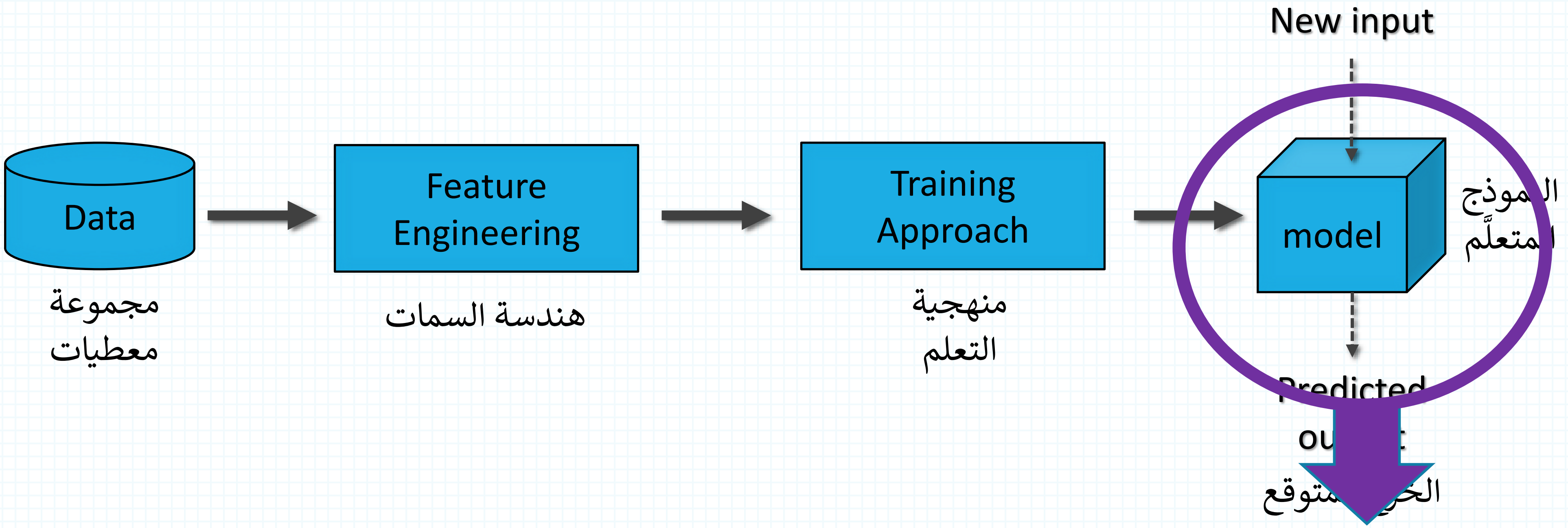
Traditional ML Pipeline



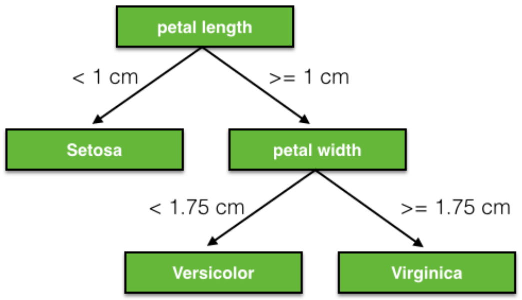
Traditional ML Pipeline



Traditional ML Pipeline

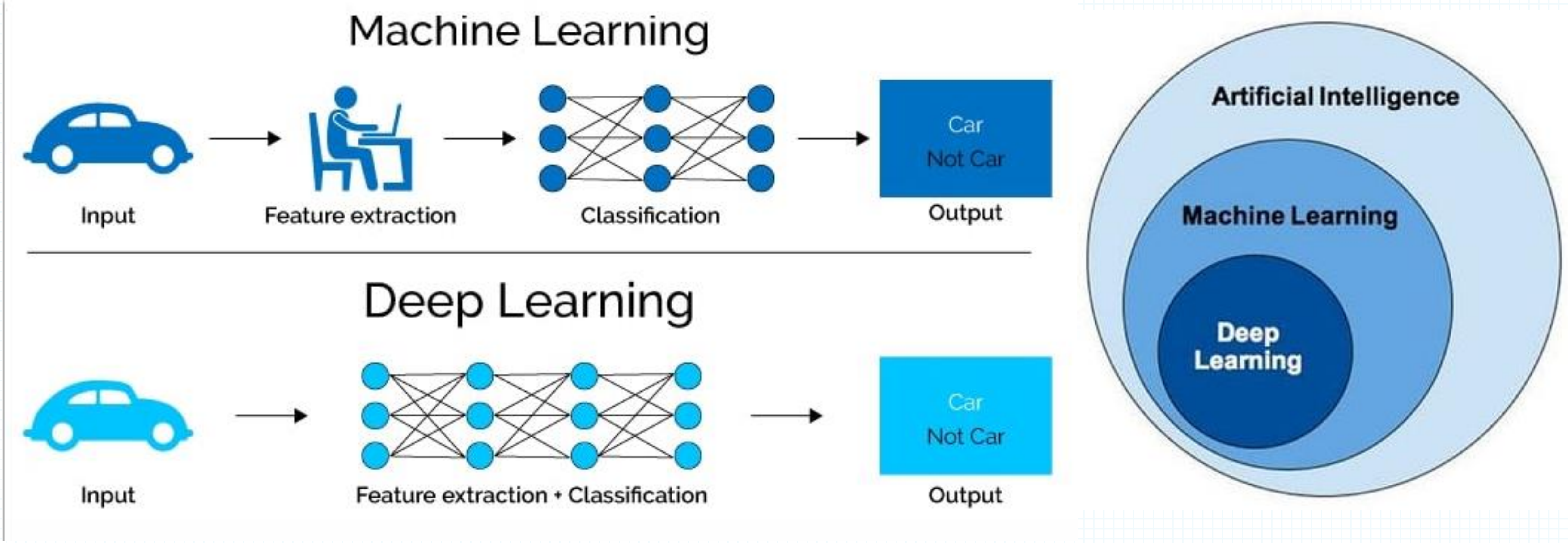


Sepal Length: 5.1 ,
Sepal Width=3.5,
Petal Length = 0.8,
Petal Width = 0.2,
Class = ???



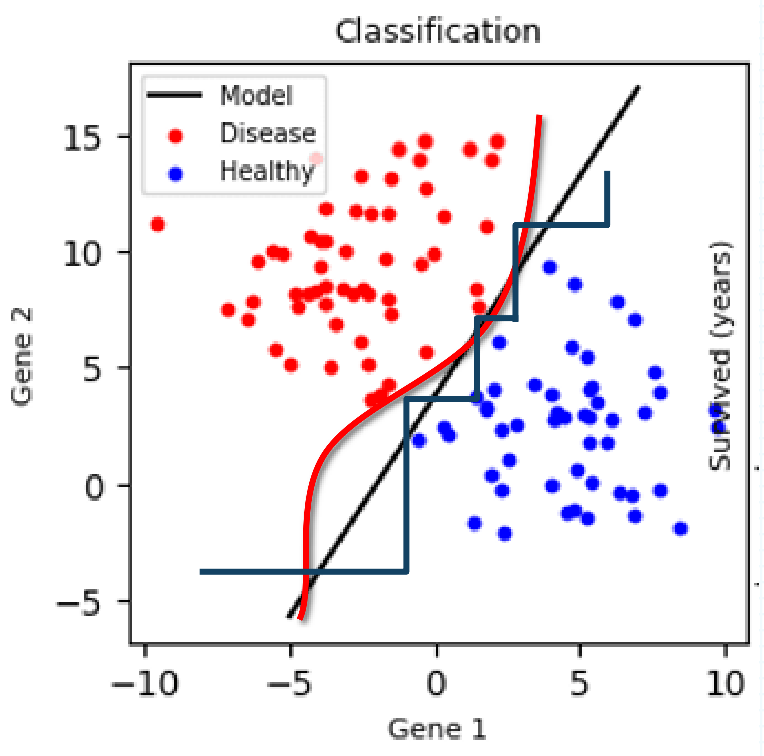
Class = Setosa

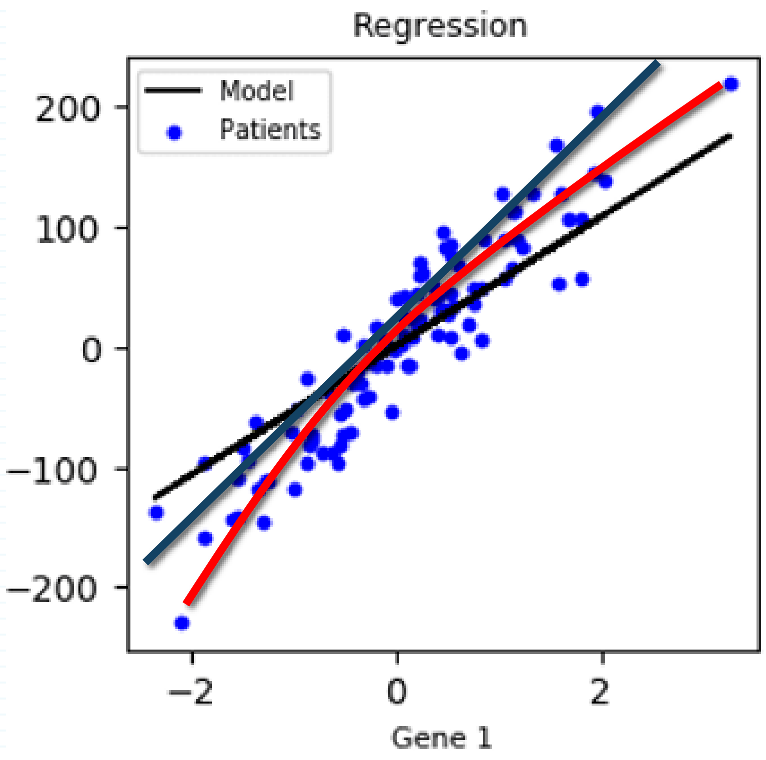
Traditional ML vs Deep Learning



Challenges in Machine Learning: Model Selection and Generalization

One of the main Challenge in ML: The vast Number of Possible ML models!

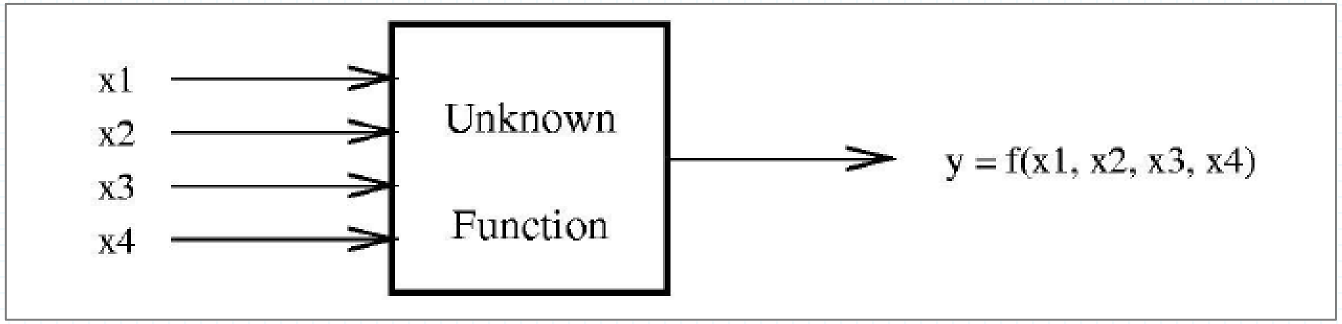




Which Dissension Boundaries is better!

One of the main Challenge in ML: The vast Number of Possible ML models!

Example



Example	x_1	x_2	x_3	x_4	y
1	0	0	1	0	0
2	0	1	0	0	0
3	0	0	1	1	1
4	1	0	0	1	1
5	0	1	1	0	0
6	1	1	0	0	0
7	0	1	0	1	0

What is the number of possible options?

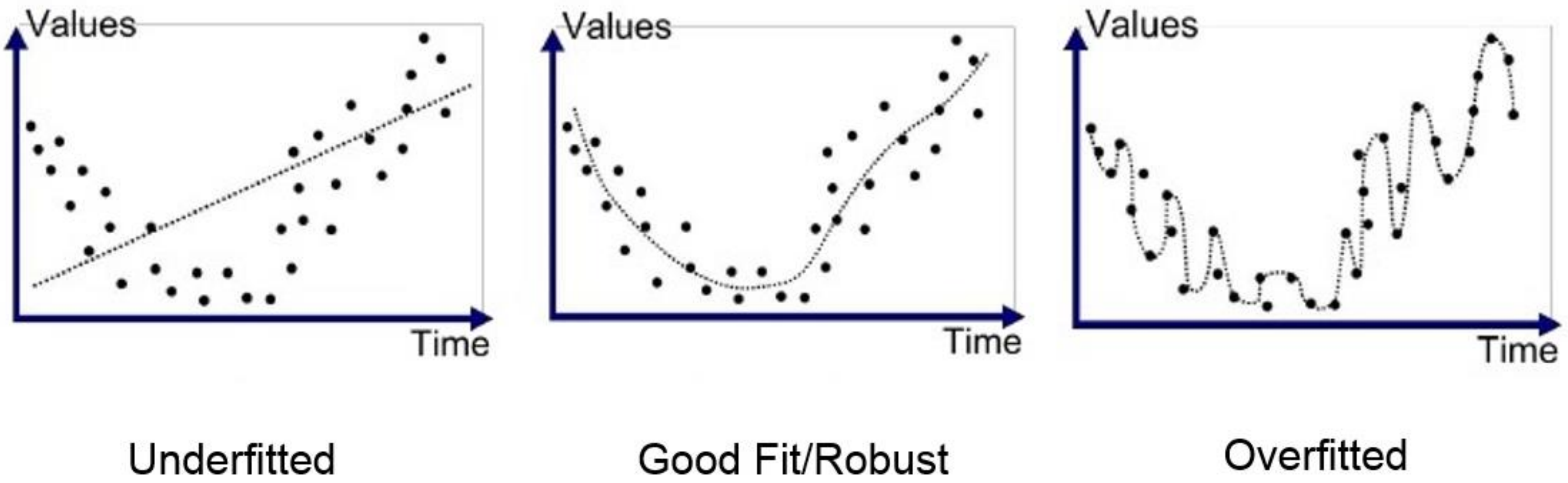
One of the main Challenge in ML: The vast Number of Possible ML models!

- 4 Boolean features
  - 2x2x2x2=16 options
  - Number of possible functions: 2^{16}
- We know 7 examples:
  - Number of possible functions: 2^9

So.. Which Model is better?

x_1	x_2	x_3	x_4	y
0	0	0	0	?
0	0	0	1	?
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	?
1	0	0	0	?
1	0	0	1	1
1	0	1	0	?
1	0	1	1	?
1	1	0	0	0
1	1	0	1	?
1	1	1	0	?
1	1	1	1	?

So.. Which Model is better?

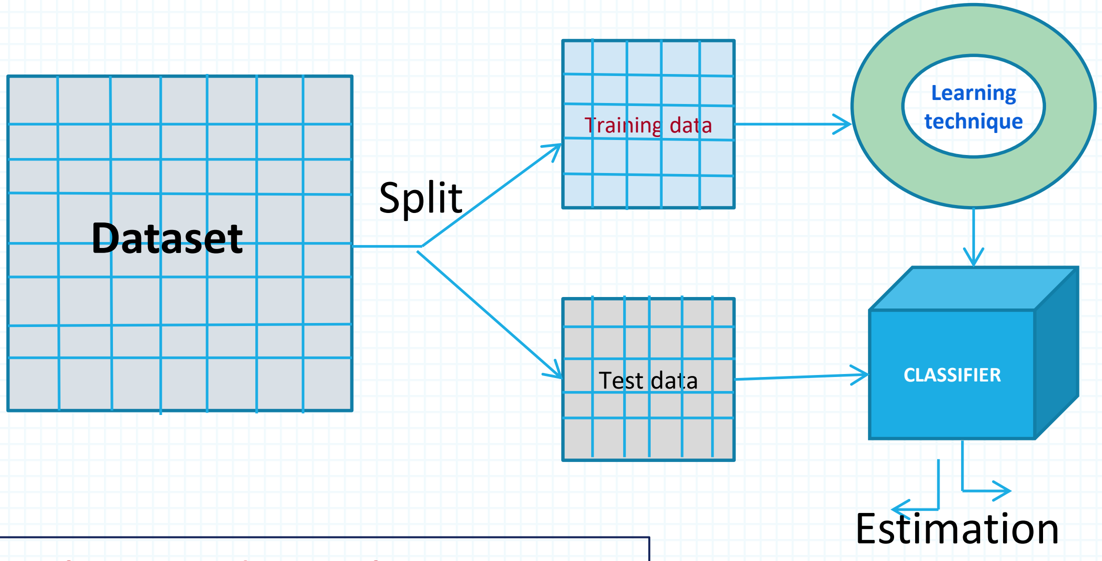


Generalization

“model's ability to adapt properly to new, previously unseen data”

How Can We Test on "Unseen" Data?
We need an Evaluation Strategy

- Split dataset into two groups
  - Training set: used to train the classifier
  - Test set: used to estimate the error rate of the trained classifier.

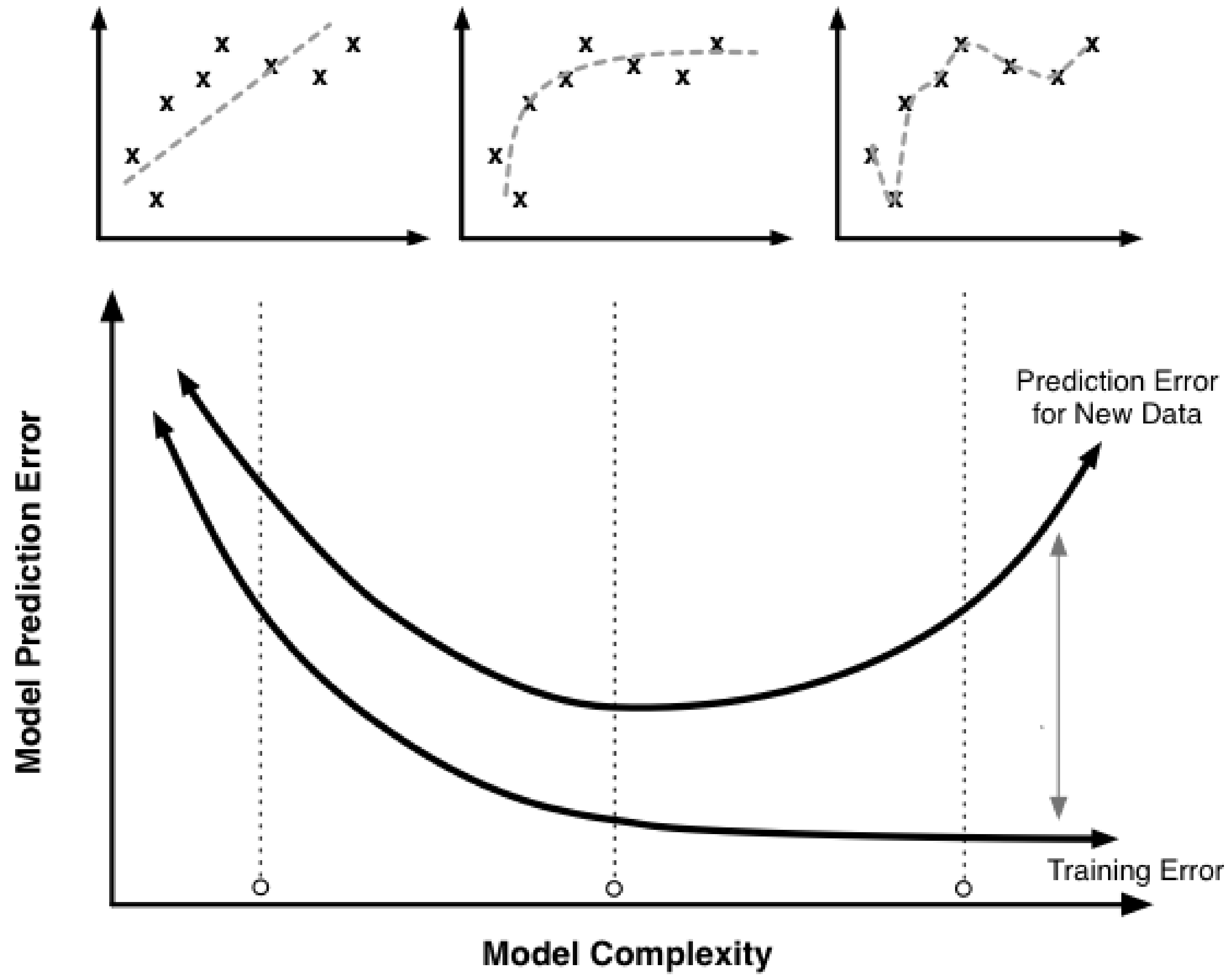


Note: This setup isn't perfect because I still see the test data when choosing the model, which can bias the results (We will discuss this point in more detail in a future lecture).

Overfitting vs. Underfitting

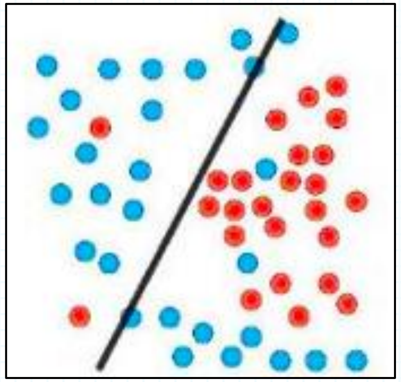
- **Overfitting:** The model learns noise and specific patterns in the training data, leading to poor performance on new data.

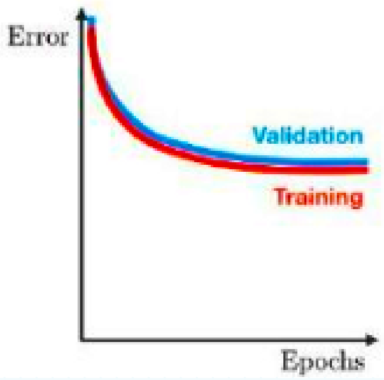
- **Underfitting:** The model is too simple and fails to capture meaningful patterns.



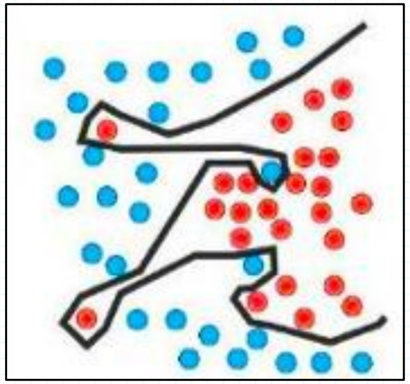
Overfitting vs. Underfitting

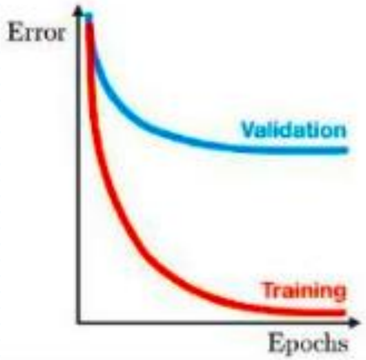
- ***Underfitting***
  - The model is too "simple" to represent all the relevant class characteristics
  - E.g., model with too few parameters produces <u>high error on the training set</u> and <u>high error on the validation set</u>





- ***Overfitting***
  - The model is too "complex" and fits irrelevant characteristics (noise) in the data
  - E.g., model with too many parameters produces <u>low error on the training set</u> and <u>high error on the validation set</u>





Next Lectures