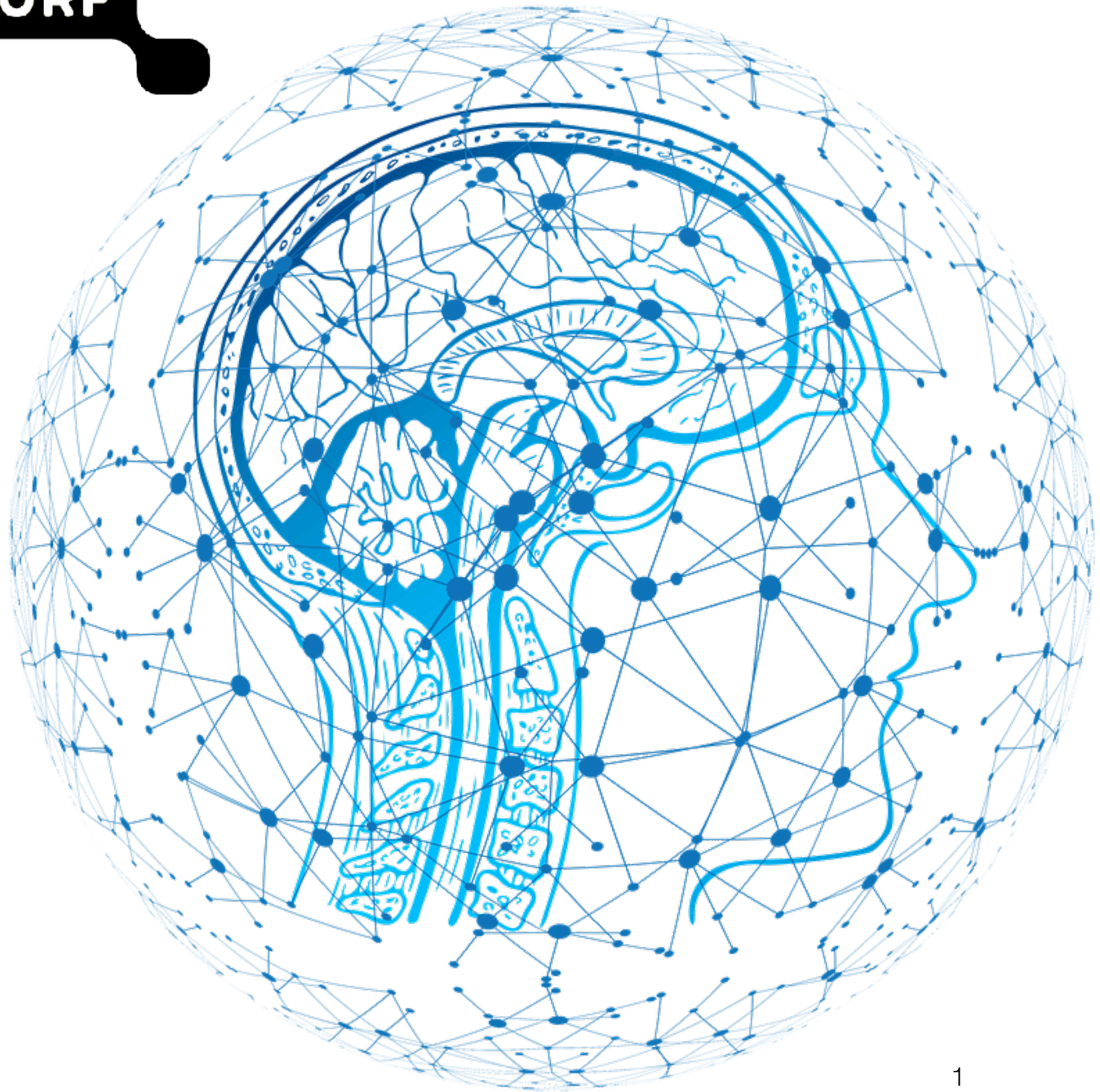
SURF



Introduction to Deep Learning

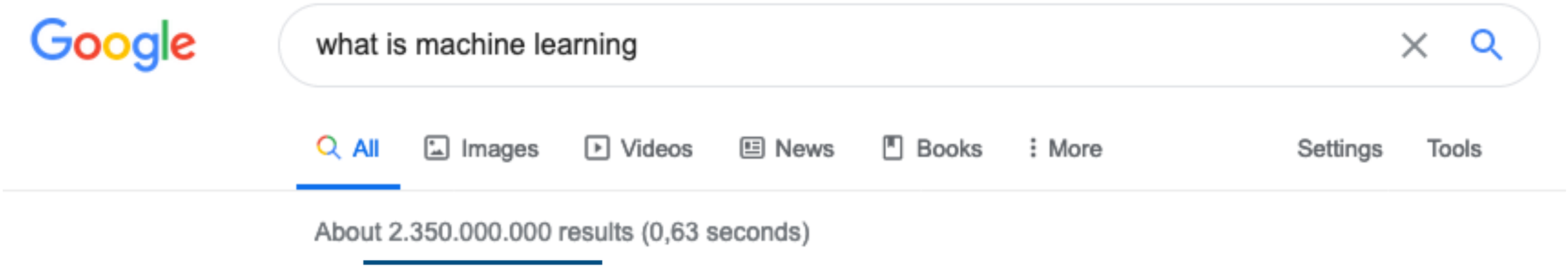
Basic Introduction

Maxwell Cai

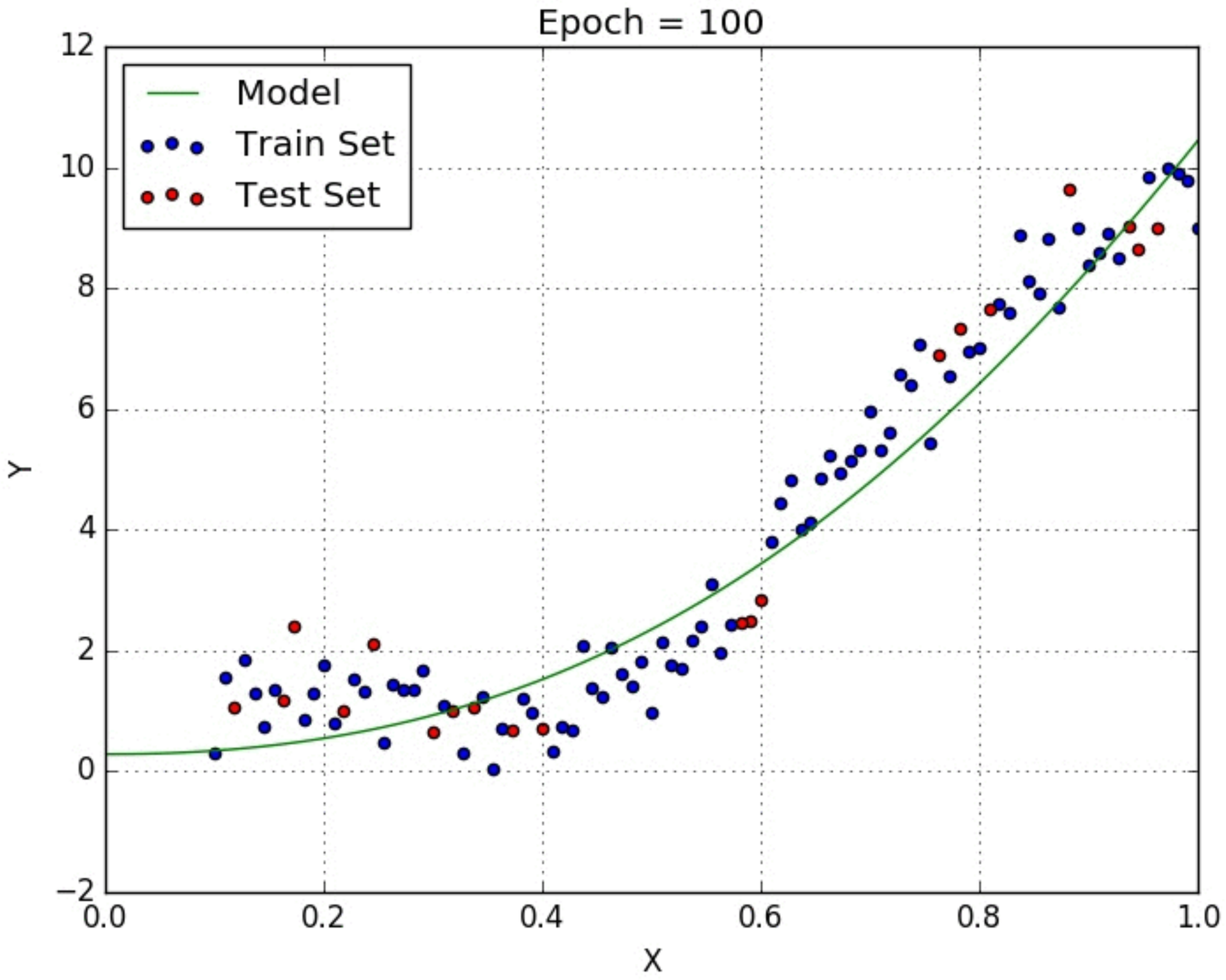
maxwellcai.com

What *is* machine learning?

Search: “what is machine learning”



What *is* machine learning?



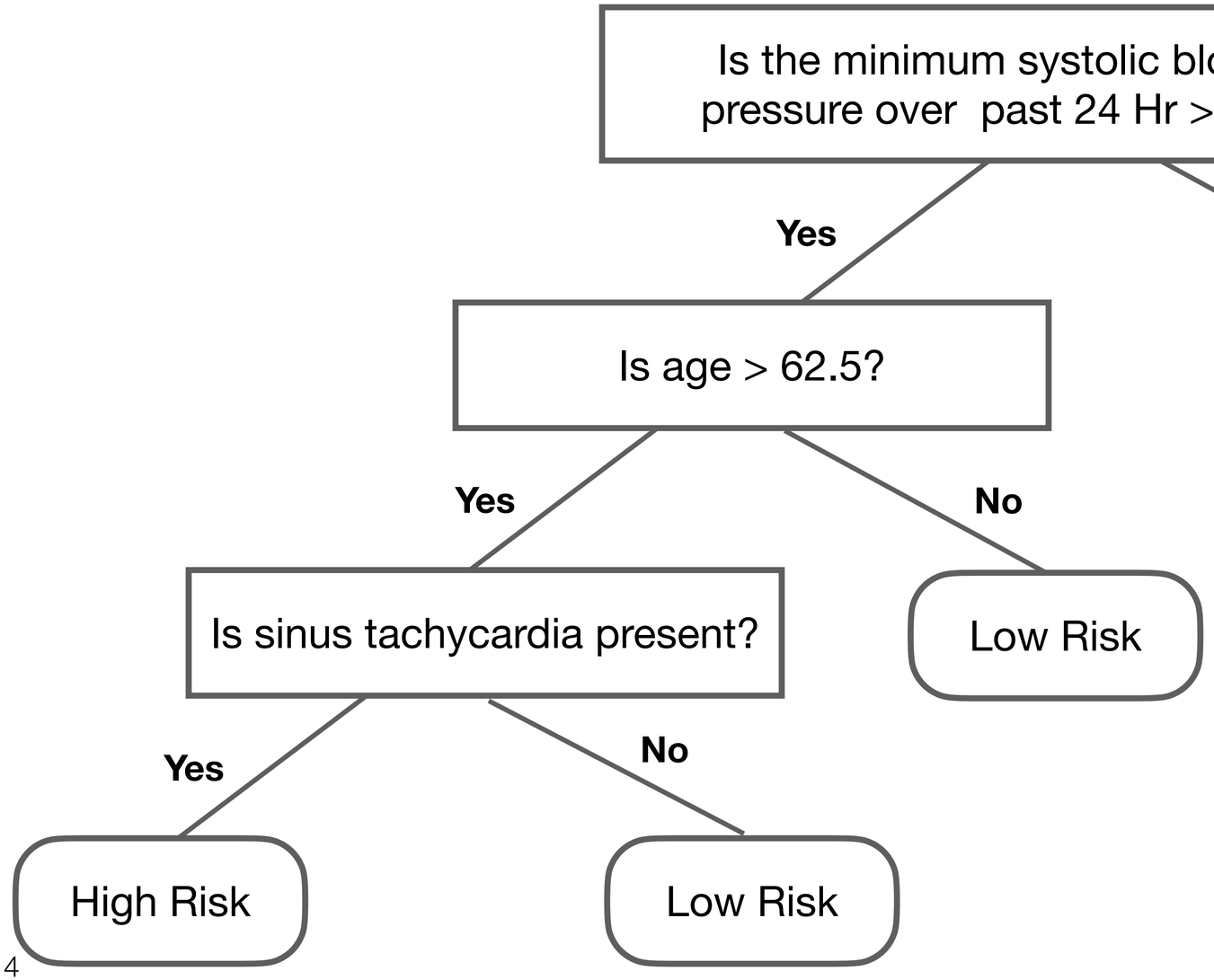
Could be as simple as curve fitting!

Typical applications in general:

- Classification
- Regression
- Dimensionality reduction
- Control

Why Machine Learning?

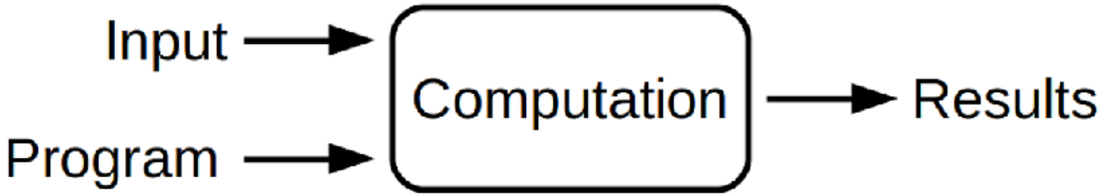
Think about a simple **decision tree**:




```
if blood_pressure > 91:
    if age > 62.5:
        if sinus_tach:
            ...
        else:
            ...
else:
    ...
```

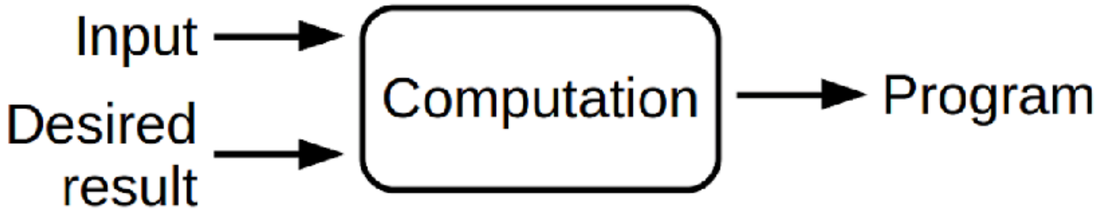
What *is* machine learning?

Traditional programming



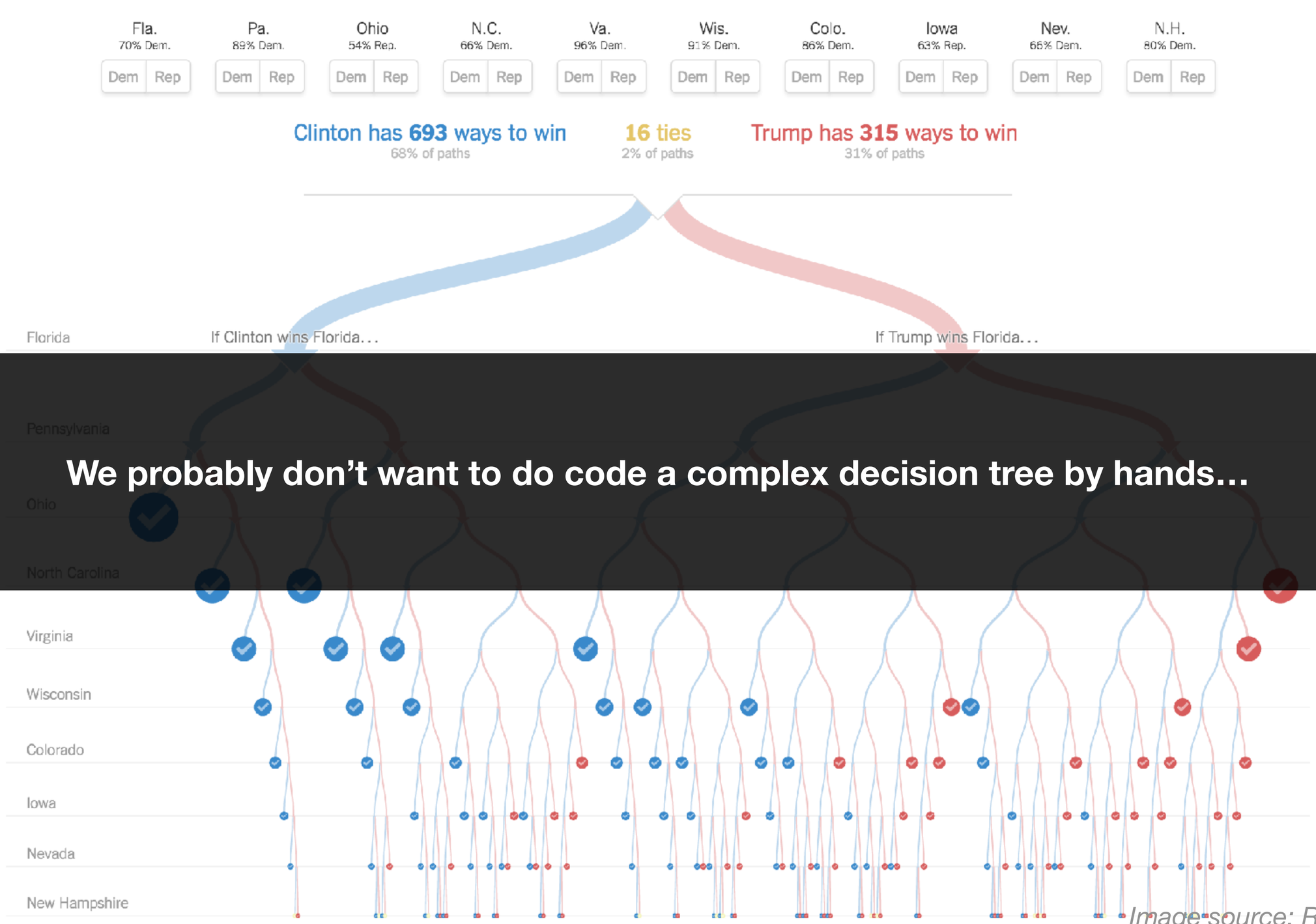
Knowledge base & rules
Expert systems
Human input

Machine learning



Features and result
'Decision' system
Limited human input

Source: Machine Learning for an Expert System to Predict Preterm Birth Risk , Woolery et al. (1994)



We probably don’t want to do code a complex decision tree by hands…

Image source: Rahul/Medium

AI vs ML vs DL

Artificial Intelligence

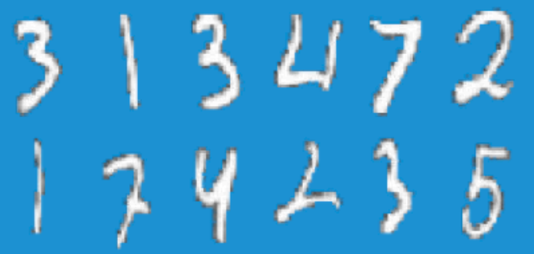
Any technique that enables computers to mimic human behavior

Machine Learning

Ability to learn without explicitly being programmed

Deep Learning

Extract patterns from data using neural networks

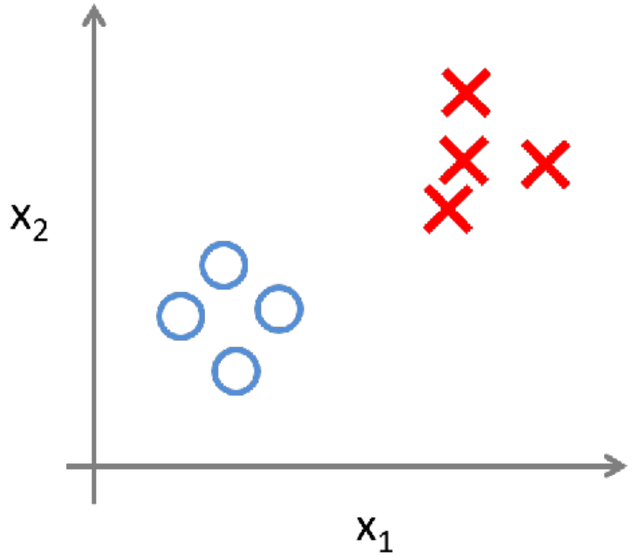


Credit: Alexander Amini/MIT

Categories of machine learning

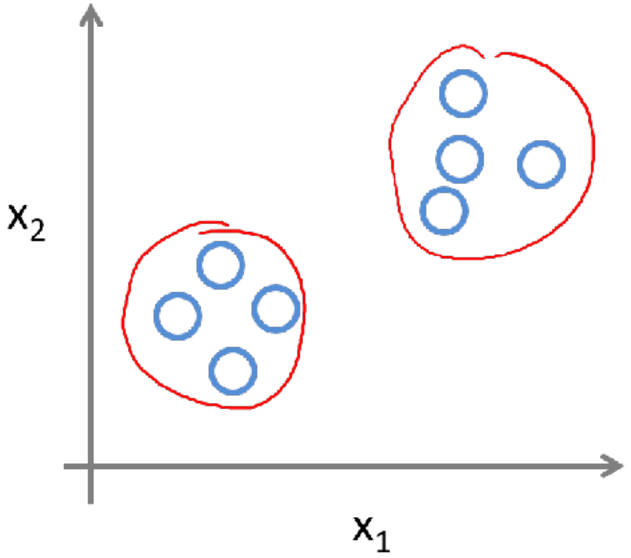
Supervised

Learn from the labels



Unsupervised

Detect patterns in the data



Reinforcement

Learn from mistakes

Categories of machine learning

Supervised

Classification
- Naive bayes
- Support vector machine
- Decision tree
- Random forest
- K-Nearest Neighbor
- Logistic regression

Regression
- Linear
- Generalized

Unsupervised

Clustering
- K-means
- K-medoids

Dimensionality reduction
- PCA
- SVD

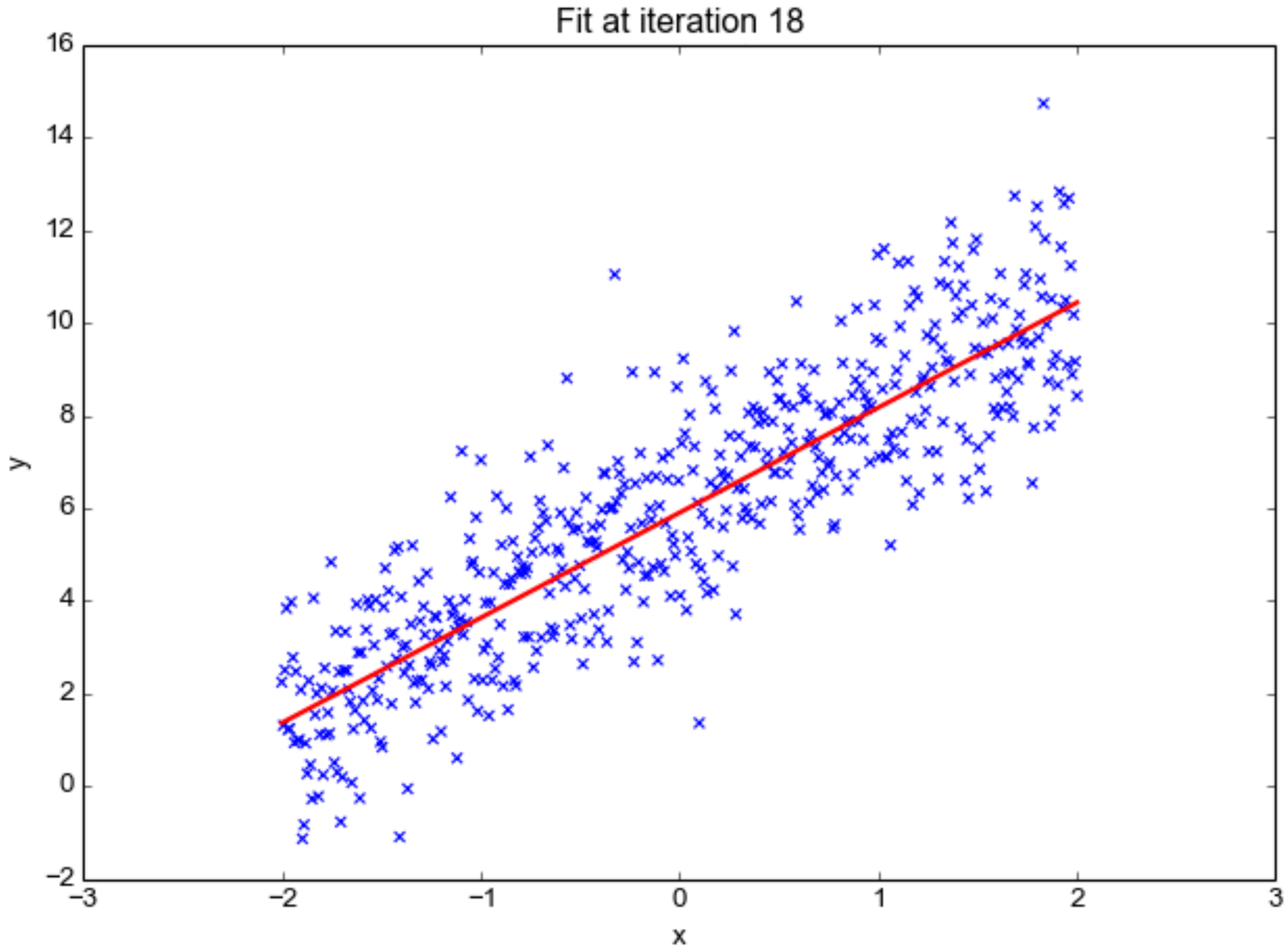
Reinforcement

Discrete
- Markov decision process
- Deep Q Network
- A2C
- A3C

Continuous
- DDPG
- NAF

… and many more…This list is incomplete!

Linear Regression



Iteratively Optimizing the parameters in a linear function

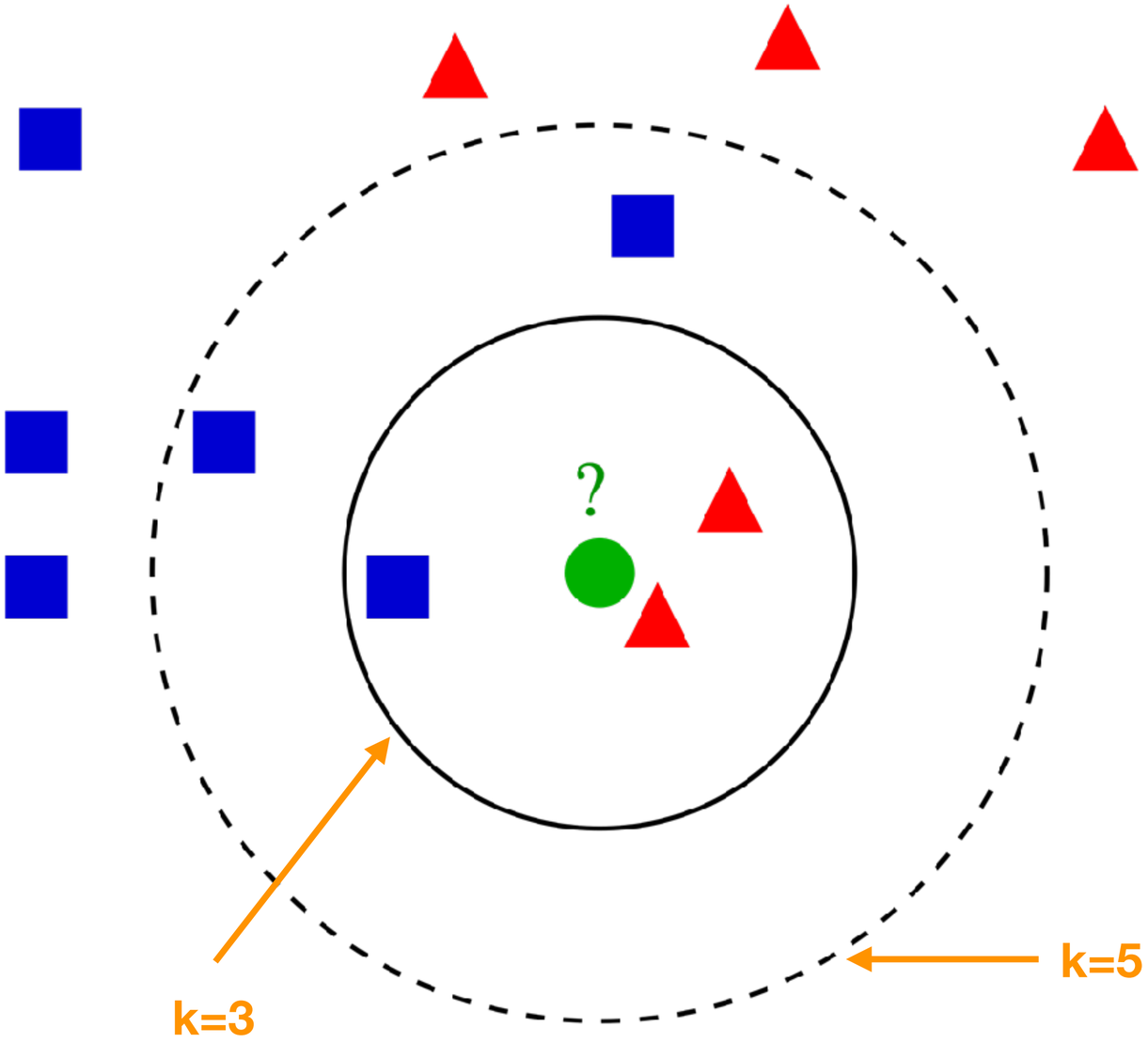
$$\hat{y} \equiv f(x) = wx + b$$

Such that

$$L(\hat{y}, y)$$

Reaches a minimum value.

Classification: *k*-Nearest Neighbor

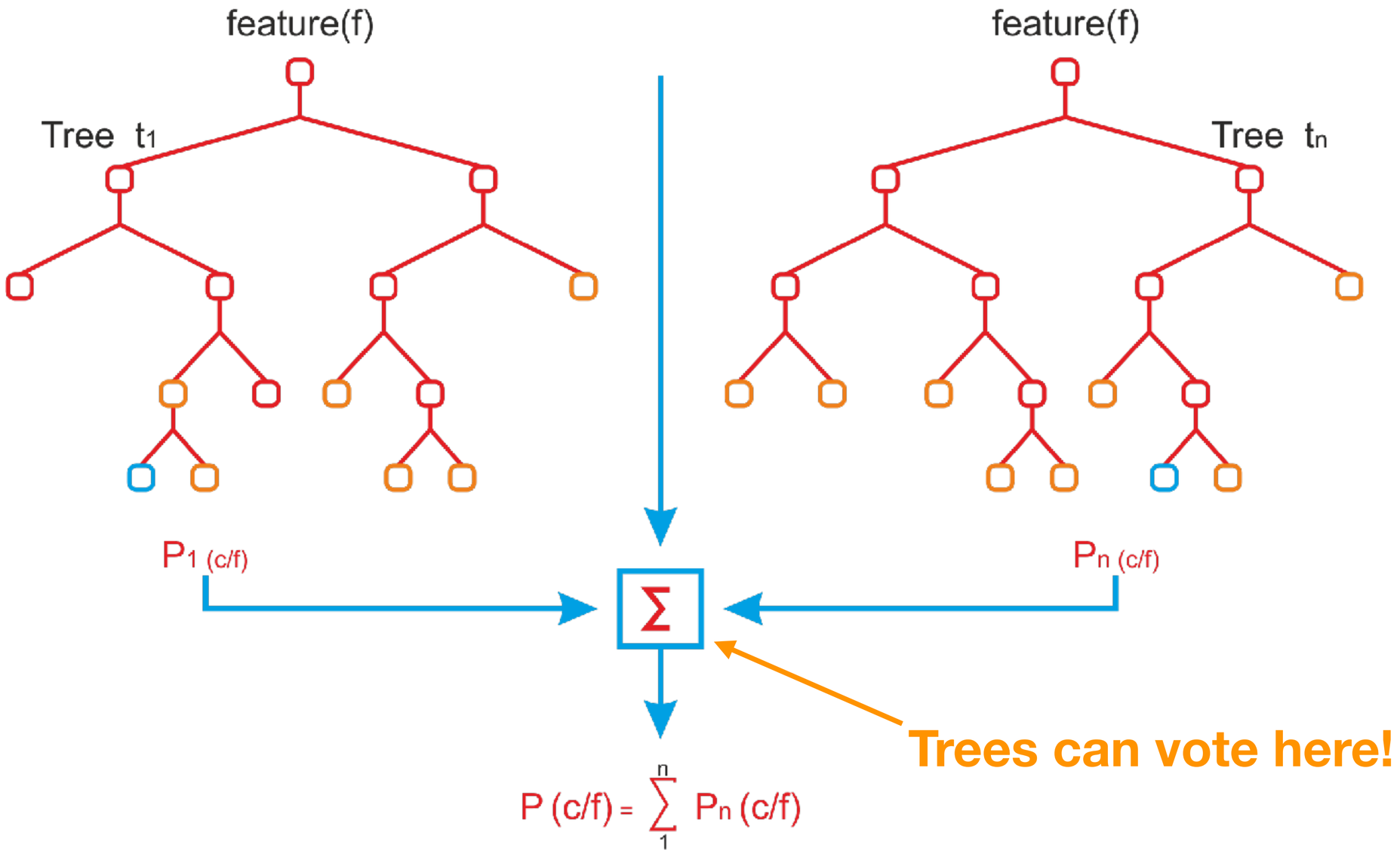


… is a **classification** algorithm

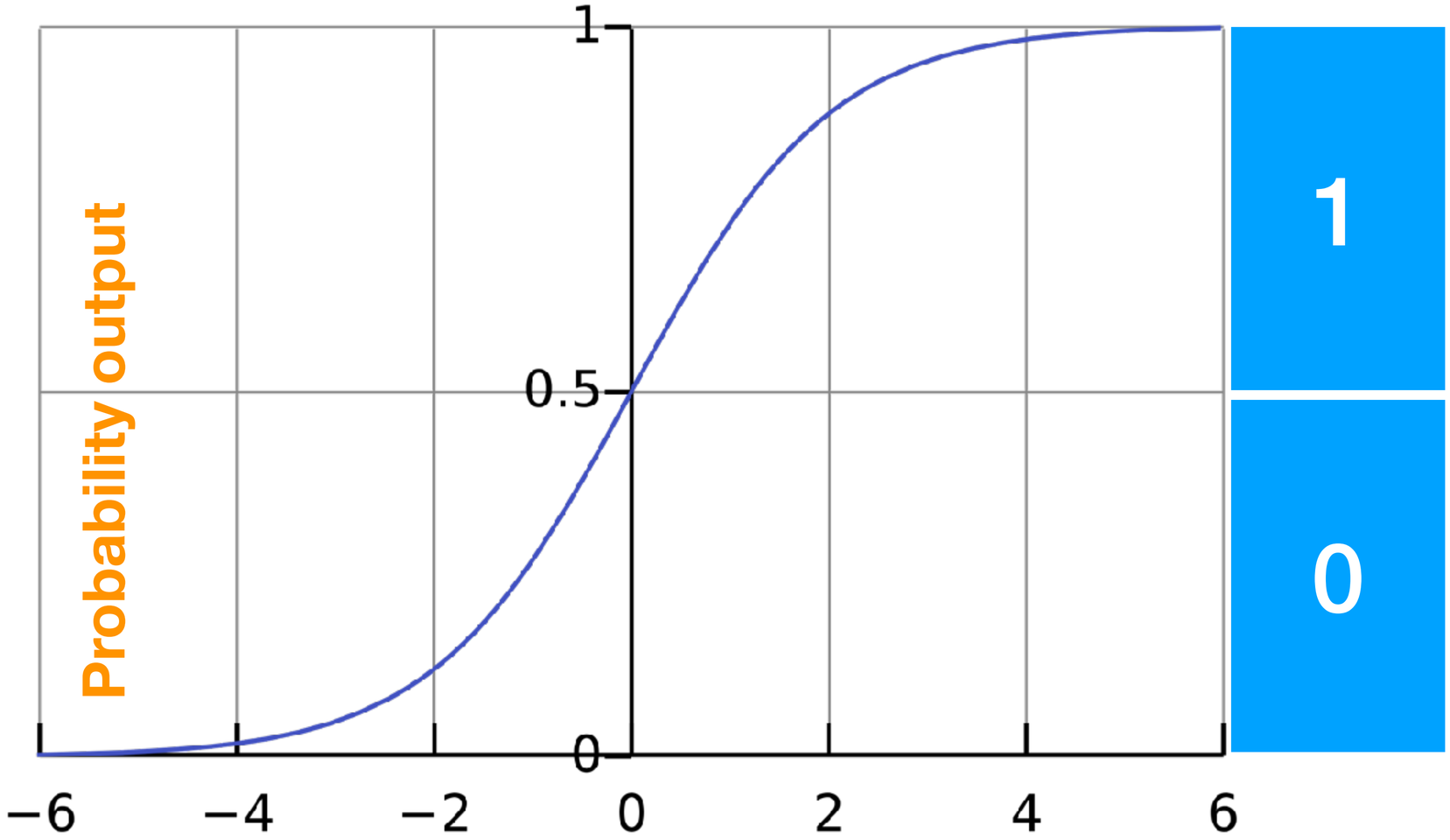
… based on a **voting** process

Image source: Wikipedia

Classification: Decision Tree → Random Forest



Classification: Logistic Regression



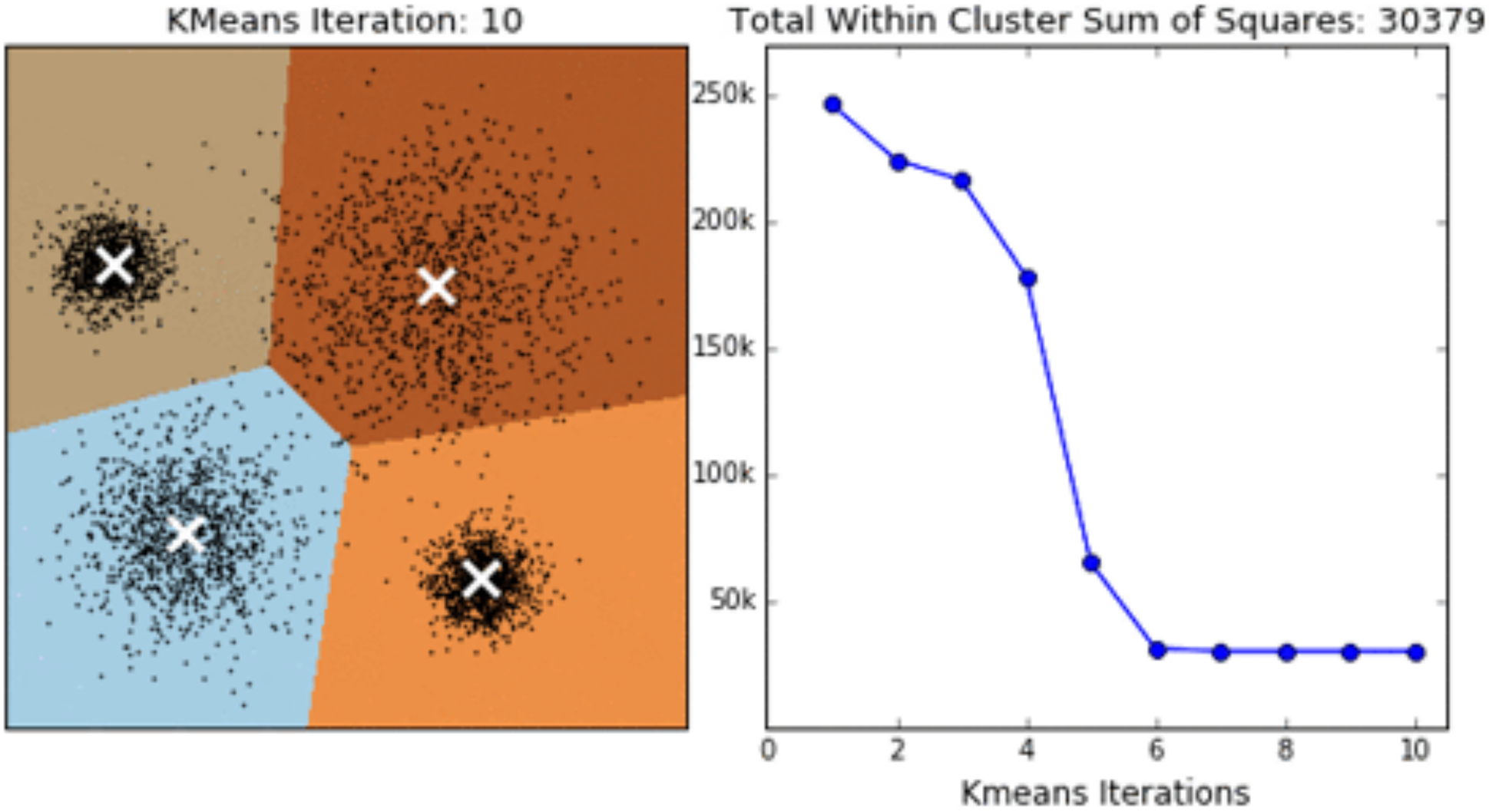
Logistic sigmoid function

$$\phi(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{x}}}$$

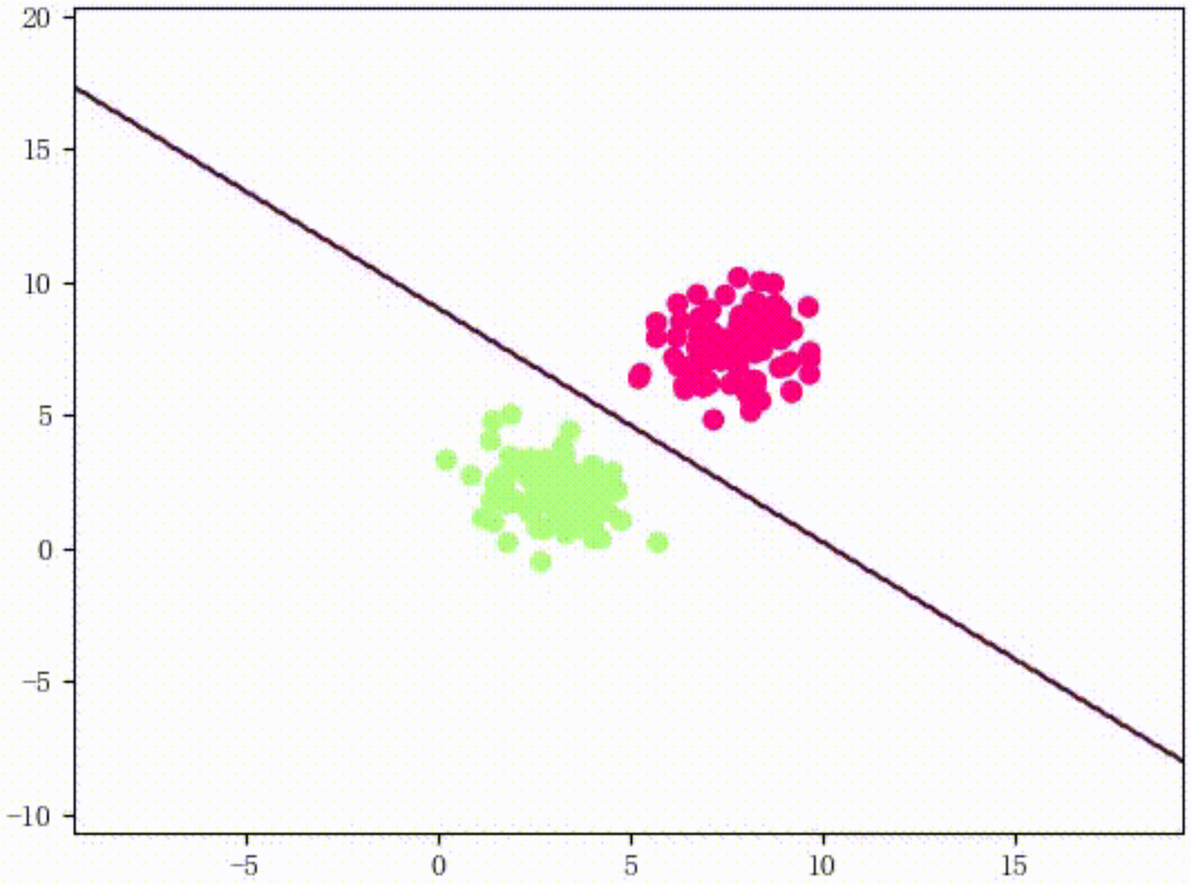
… is a binary **classification** algorithm, not a regression algorithm

… Gives the **probability** of each class

K-means clustering



Support Vector Machine



… is a **classification** algorithm

… does not provide probabilities, but only output class identities.

… aims to find an optimal way (perpendicular to the **support vector**) to separate different classes

Animation source: GitHub.com/carefree0910

Boser et al. 1992; Cortes & Vapnik 1995

What is the machine actually *learning*?

Machine learning is an optimisation process

Learning: Optimizing the loss function

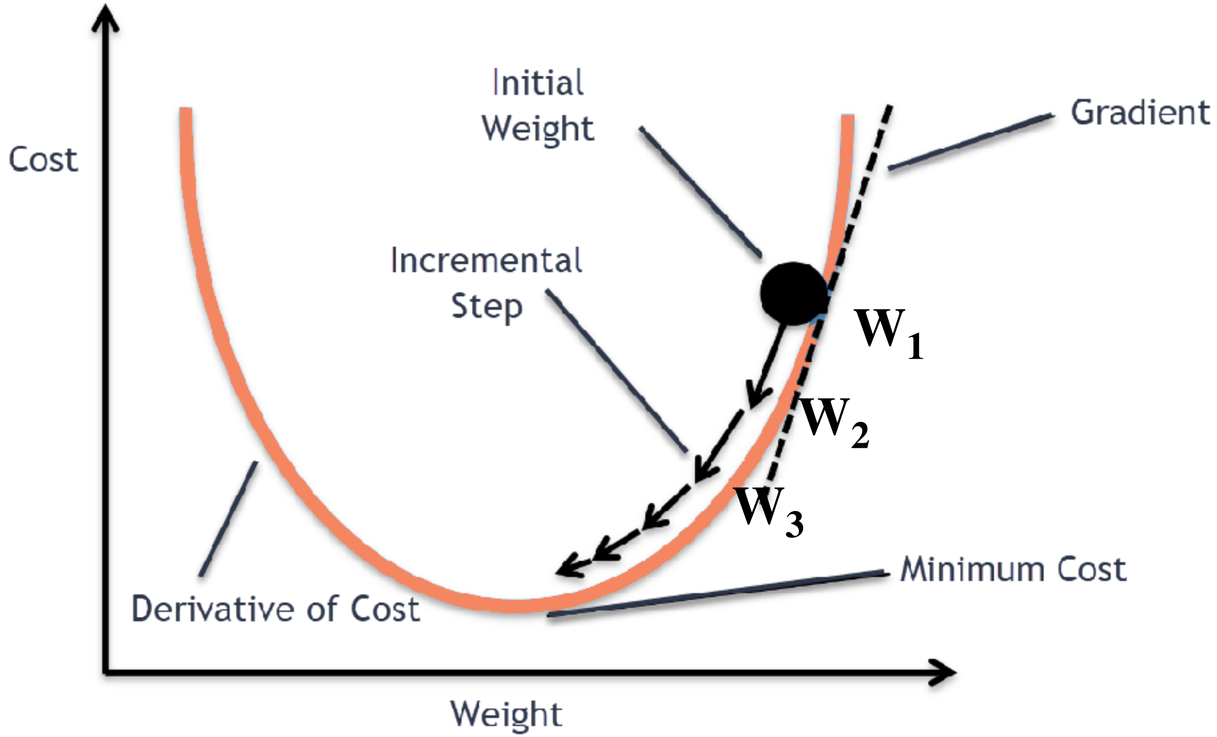
Objective: minimize the **difference** between predicted value the actual value

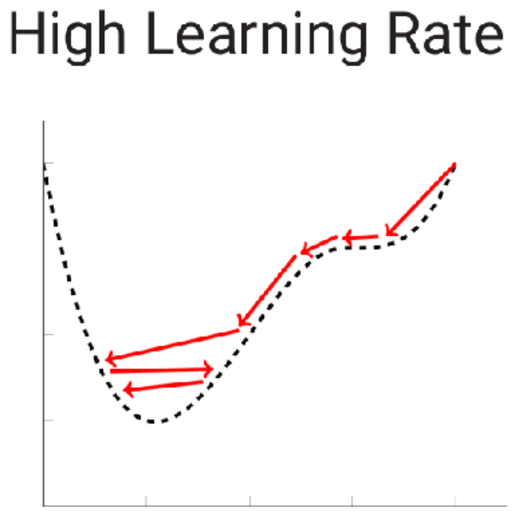
$$\mathscr{L}(\mathbf{W}, b) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2$$

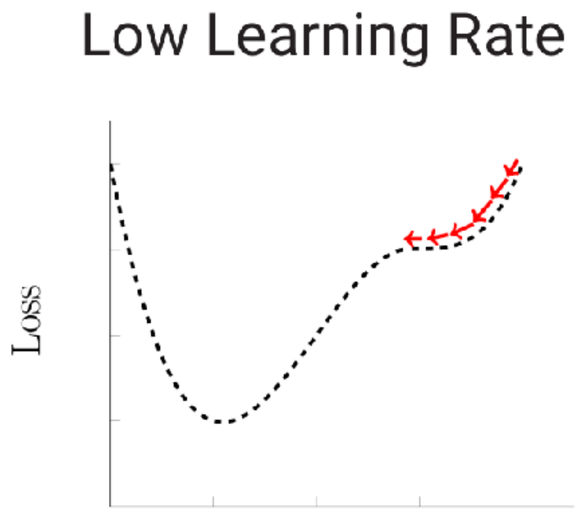
$$\nabla\mathscr{L}(\mathbf{W}, b)$$

$$\mathbf{W}_{j+1} = W_j - \alpha\nabla\mathscr{L}(\mathbf{W}_j, b)$$

Learning rate







A few more terms…

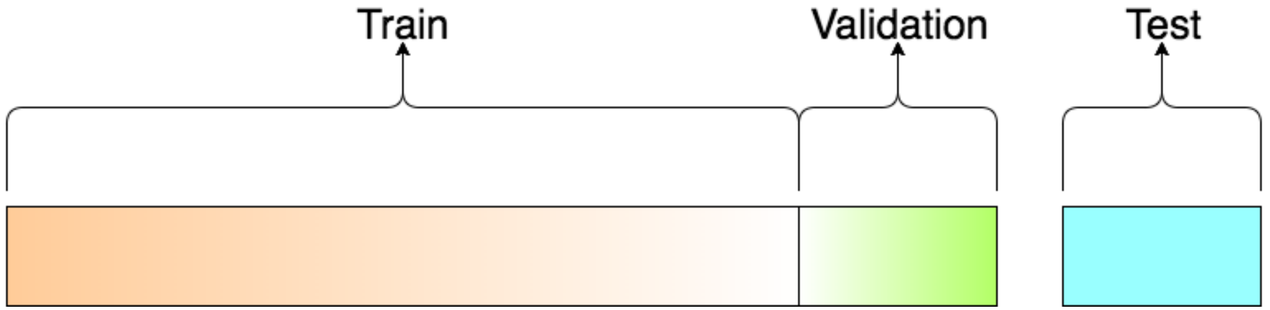
Weights

Bias

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

Estimate

Input features



Evaluation with the **mean squared error**

$$\text{MSE}(\mathbf{W}, b) = \frac{1}{N} \sum_{i=1}^{N} \left(y_i^{(\text{test})} - \hat{y}_i^{(\text{test})} \right)^2$$

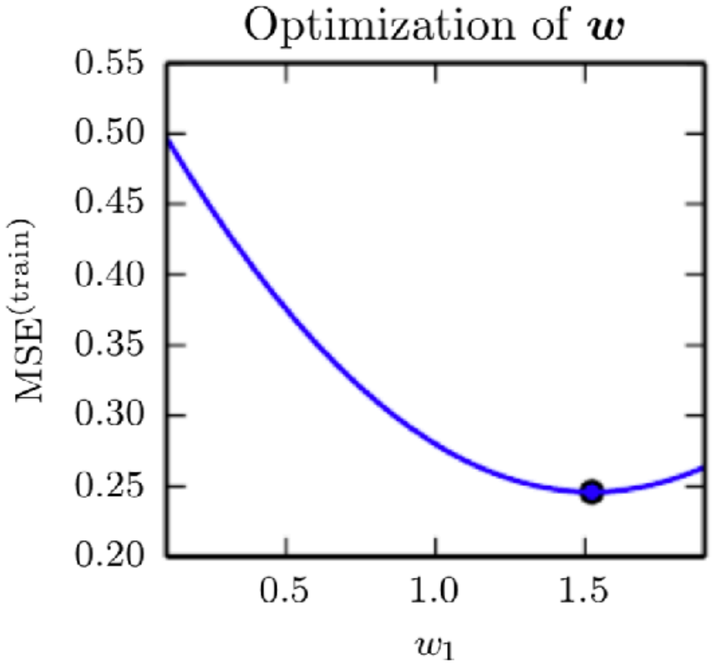
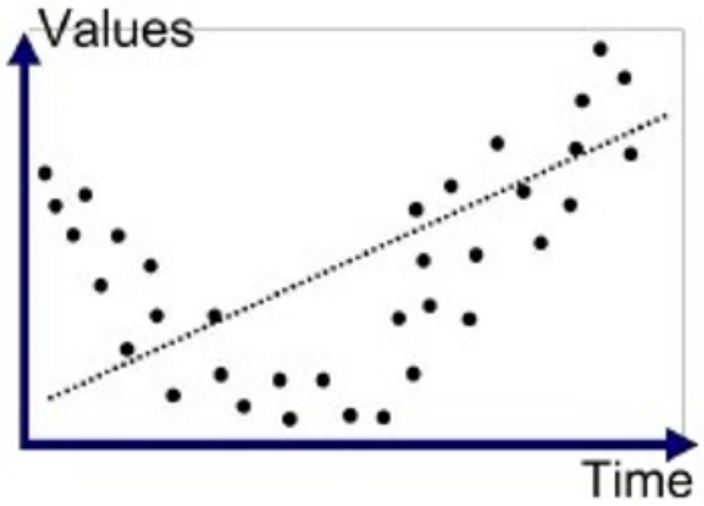
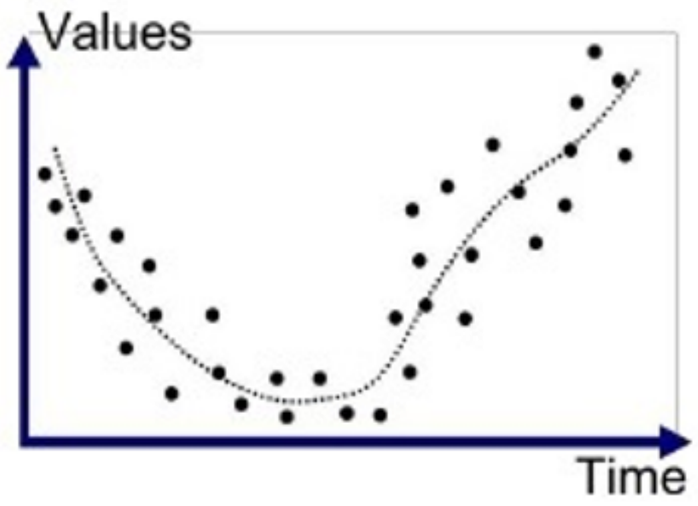


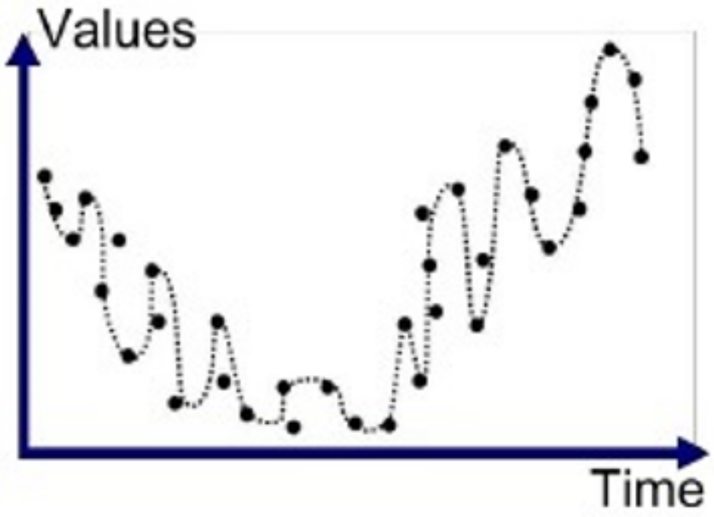
Image credits: Deep Learning (Goodfellow, Bengio & Courville); Tarang Shah

The bias-variance tradeoff

The lower the losses, the better? **Not really.**

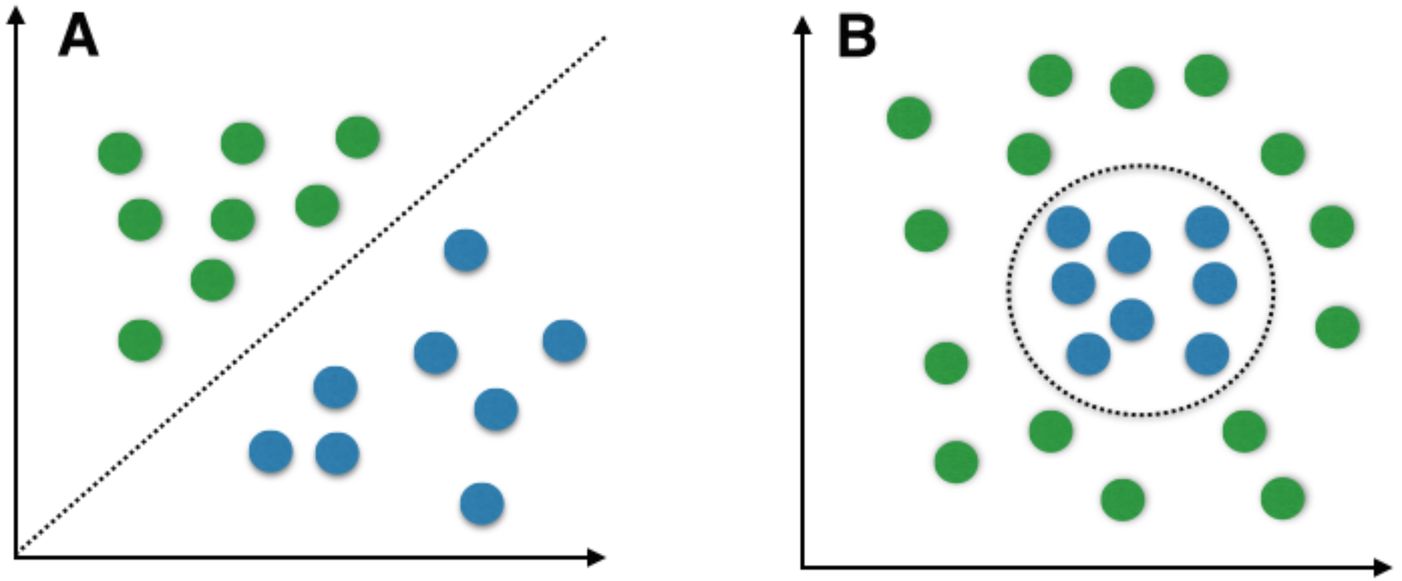


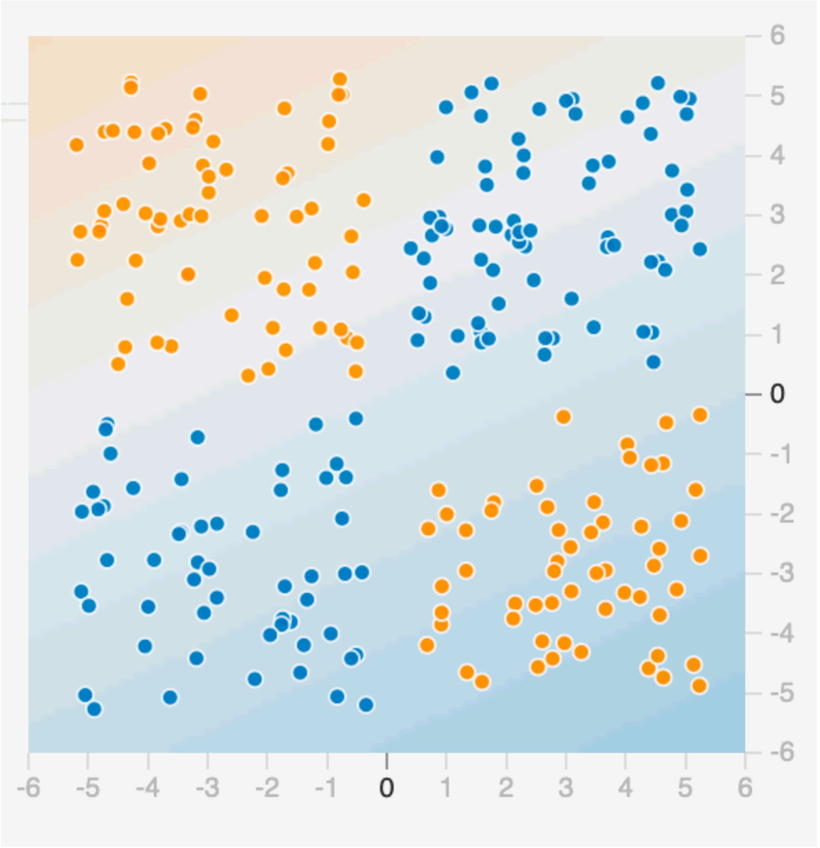




What if the problem is more complicated?

What if the problem is not so simple?





Possible solutions:

- map into another space (kernel trick)
- Add more layers (deep learning)

Kernel Trick