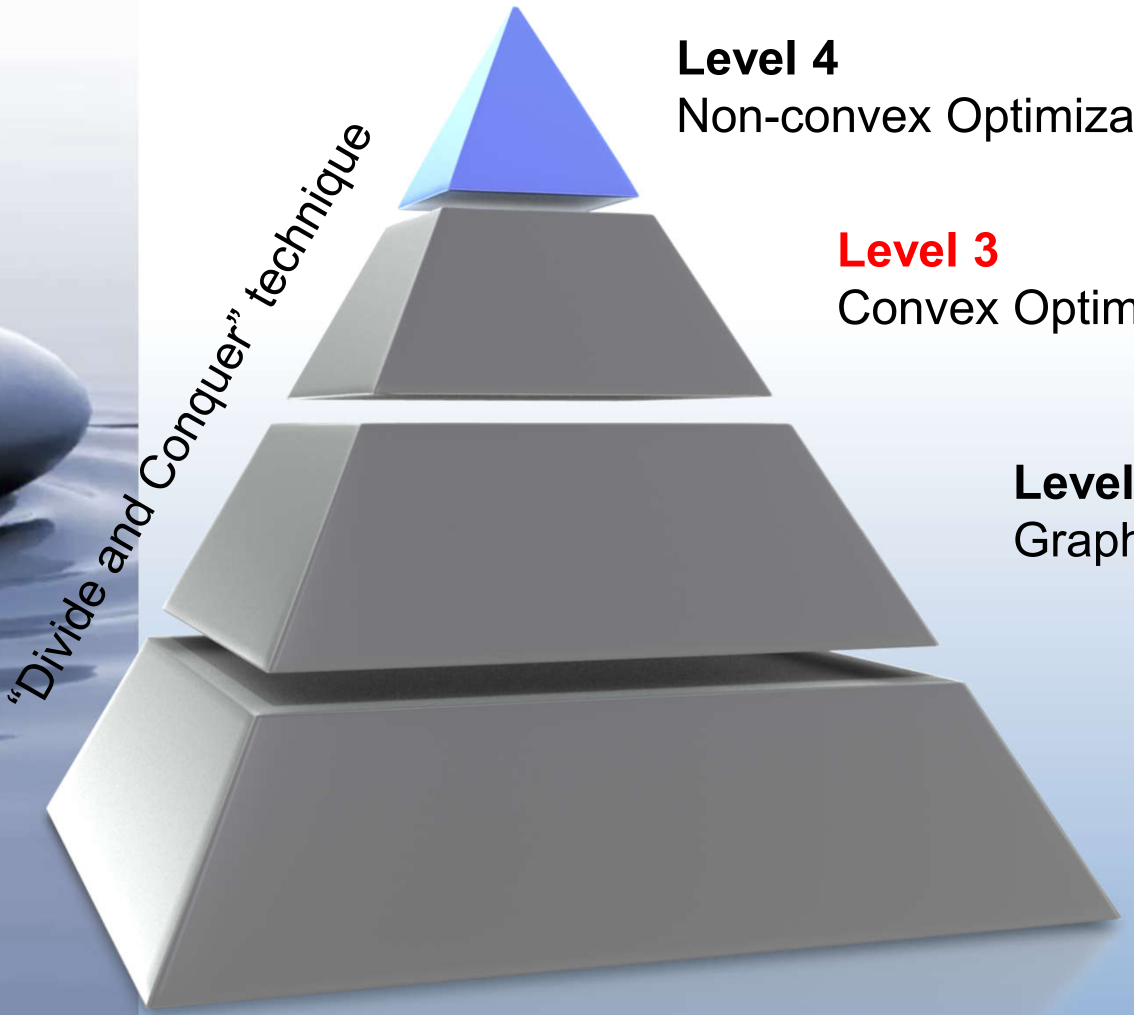
Artificial Intelligent: Introduction

Presenter: Dr. Ha Viet Uyen Synh.

How to solve a problem in computer science?

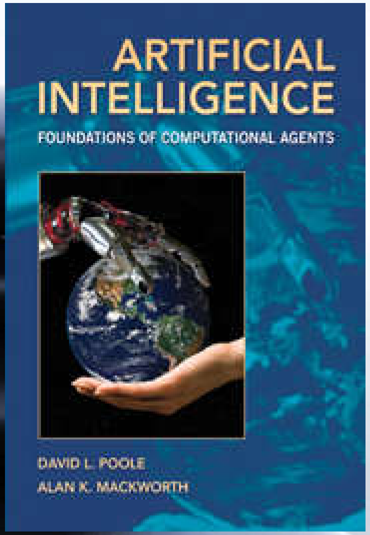


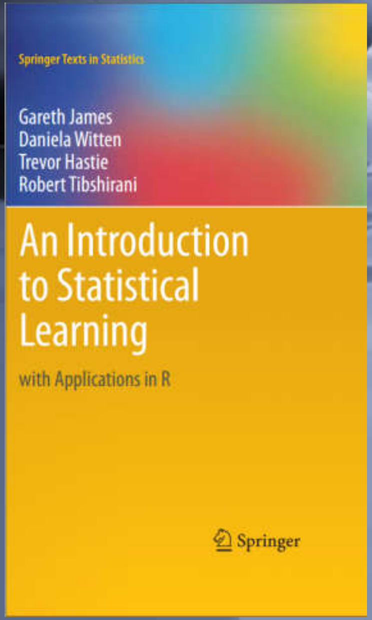
Level 4
Non-convex Optimization

Level 3
Convex Optimization

Level 2
Graphs

Level 1
“Exhausted check” method

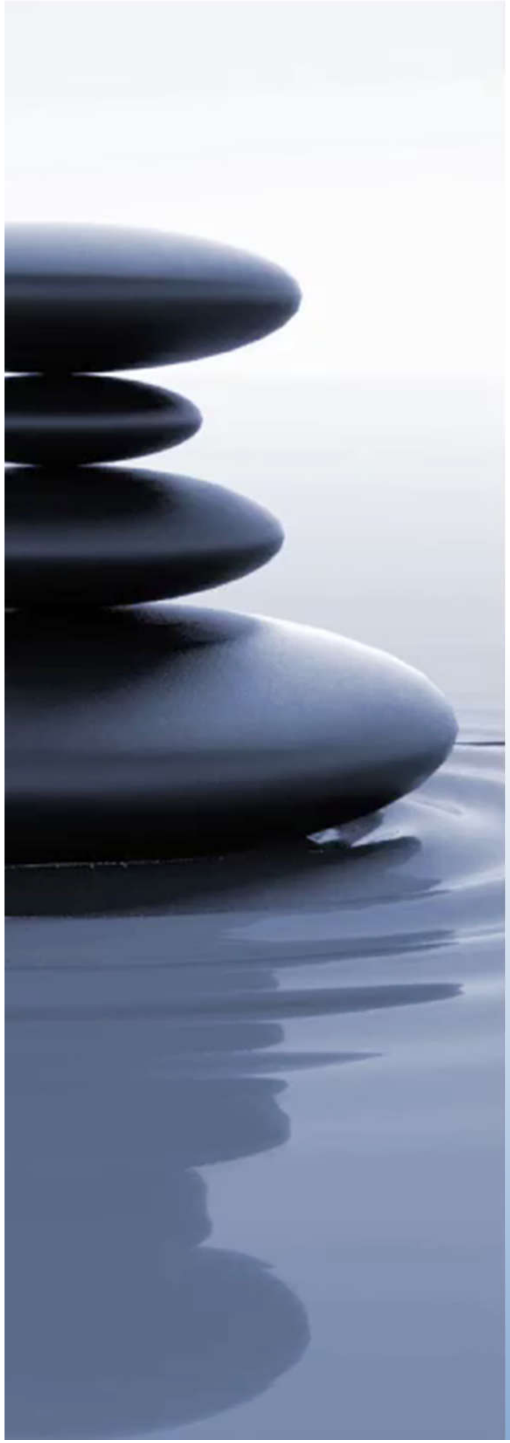




Course Materials

- Main Textbook
  - Artificial Intelligence: Foundations of Computational Agents. By Poole and Mackworth. (P&M)
  - Available electronically (free) http://artint.info/html/ArtInt.html
  - We will cover Chapters: 1, 3, 4, 5, 6, 7, 8, 9
- https://machinelearningcoban.com/2016/12/26/introduce

- Reference book
  - An Introduction to Statistical Learning with Applications in R. By Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
  - R programming language



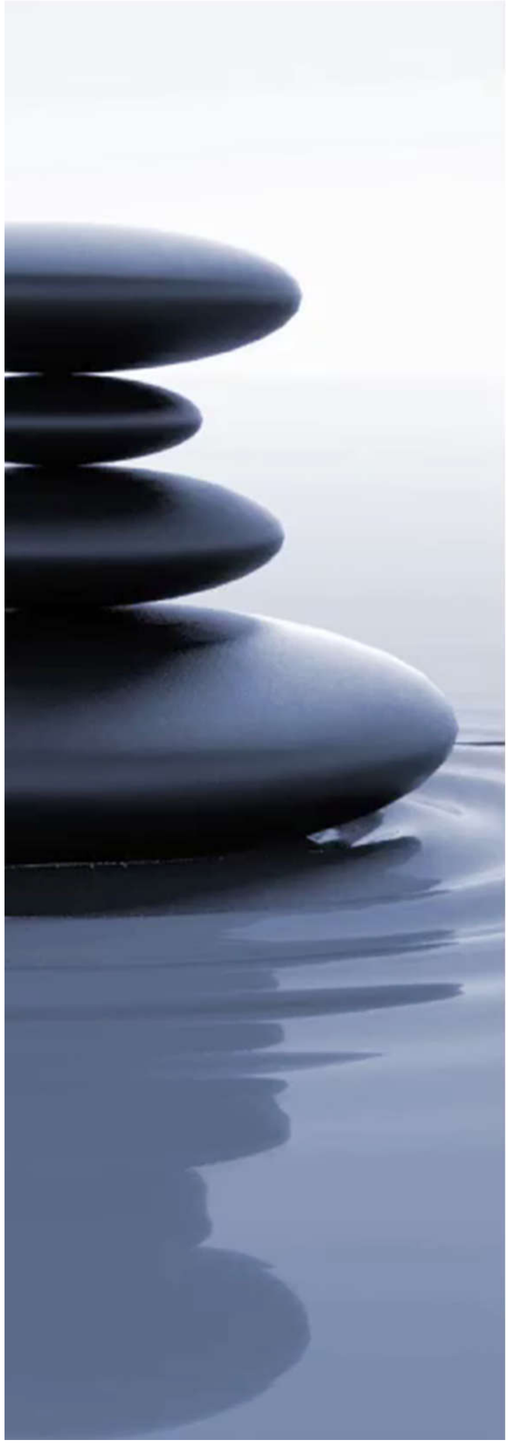
Grading

Quiz: 10%

Lab. Assignments: 20%

Midterm Test: 30%

Final Exam: 40%

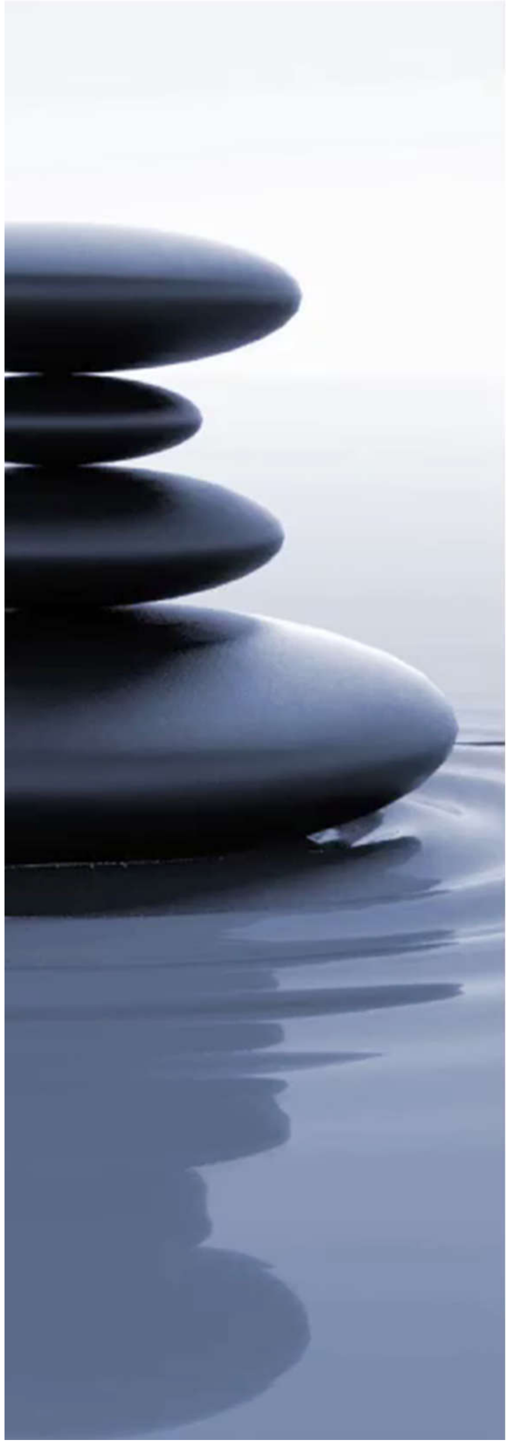


What is Machine Learning?

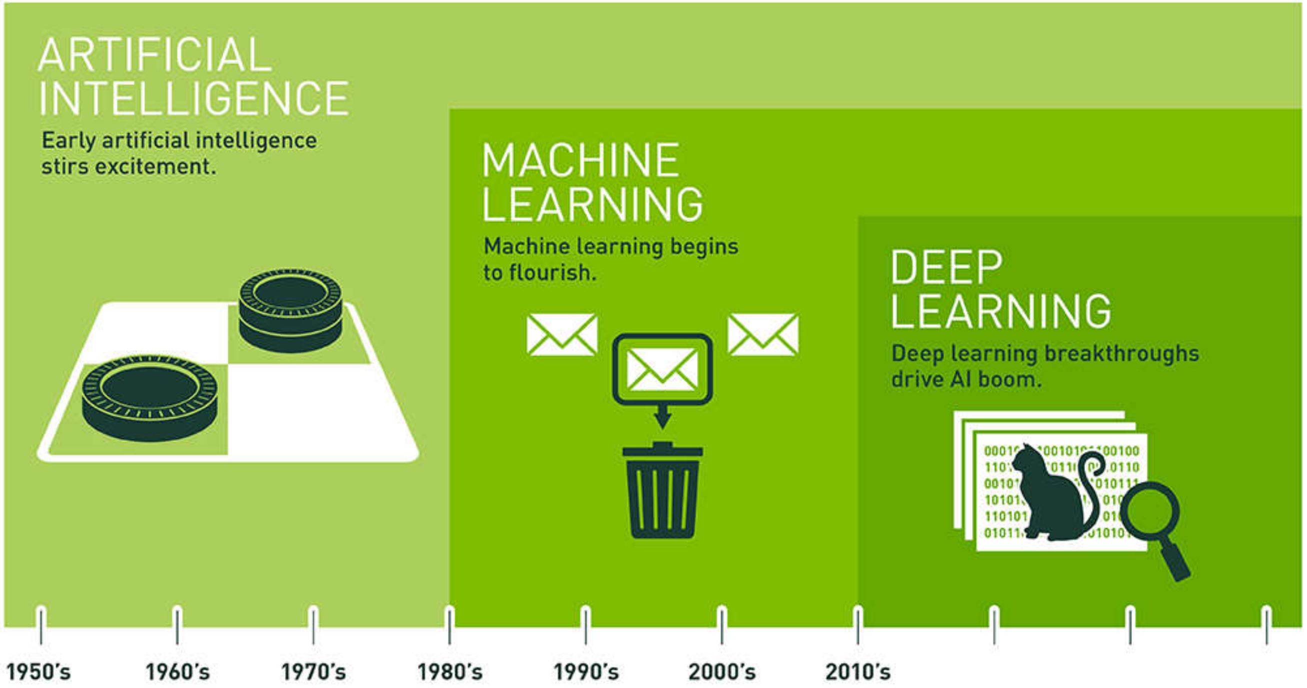
Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals: learn from experience.

Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model.

The algorithms adaptively improve their performance as the number of samples available for learning increases.



What’s the Difference Among Artificial Intelligence, Machine Learning, and Deep Learning?



ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.

MACHINE LEARNING

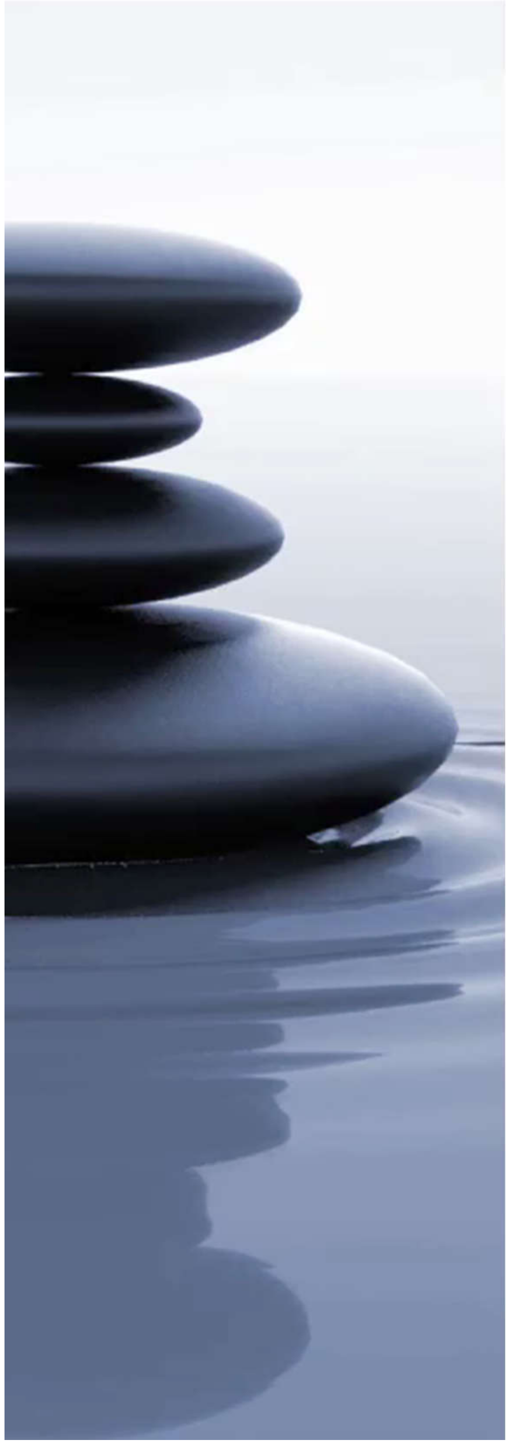
Machine learning begins to flourish.

DEEP LEARNING

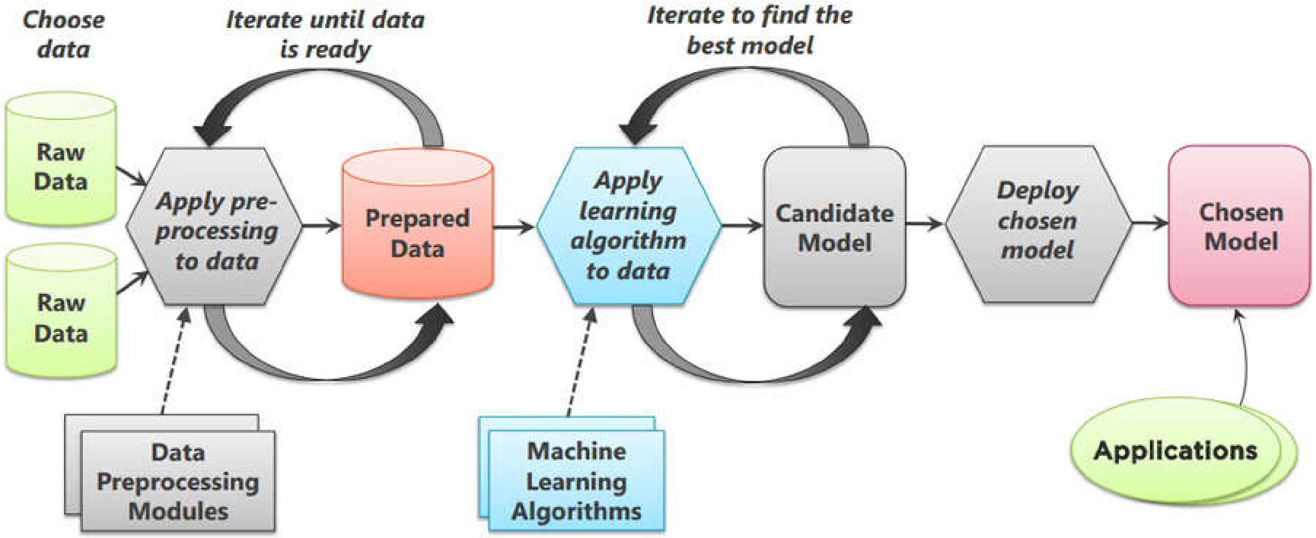
Deep learning breakthroughs drive AI boom.

Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

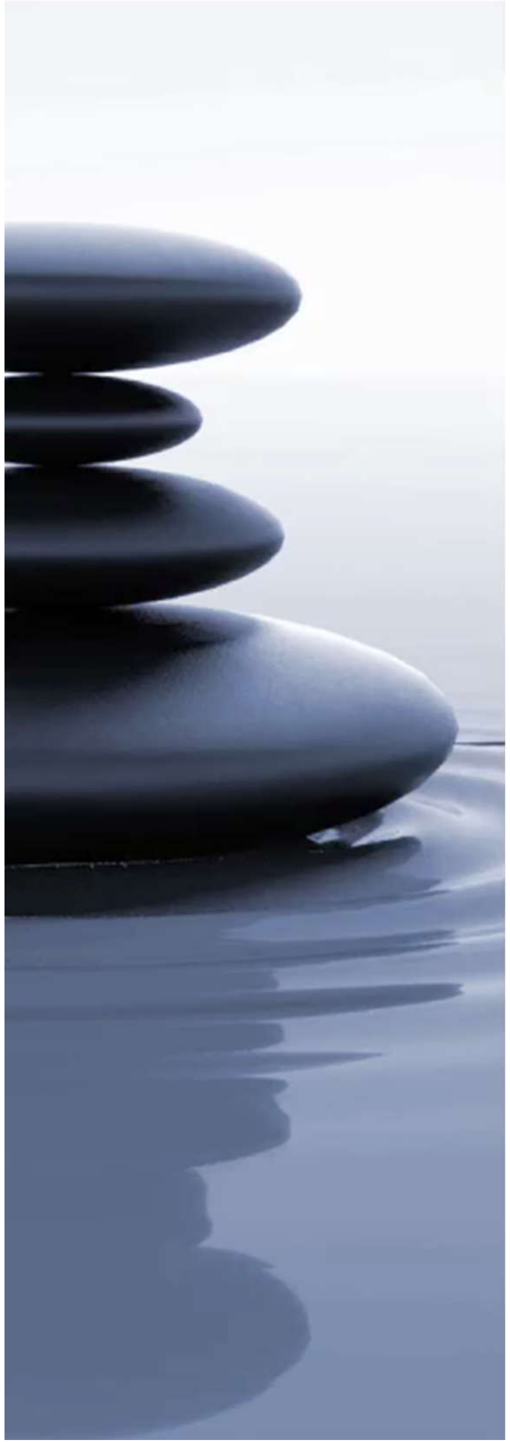
Artificial Intelligence — Human Intelligence Exhibited by Machines
Machine Learning — An Approach to Achieve Artificial Intelligence
Deep Learning — A Technique for Implementing Machine Learning



Machine Learning Process



Choose data
Iterate until data is ready
Iterate to find the best model
Raw Data
Raw Data
Apply pre-processing to data
Prepared Data
Apply learning algorithm to data
Candidate Model
Deploy chosen model
Chosen Model
Data Preprocessing Modules
Machine Learning Algorithms
Applications



Why Machine Learning Matters

With the rise in big data, machine learning has become a key technique for solving problems in areas, such as:

Computational finance, for credit scoring and algorithmic trading

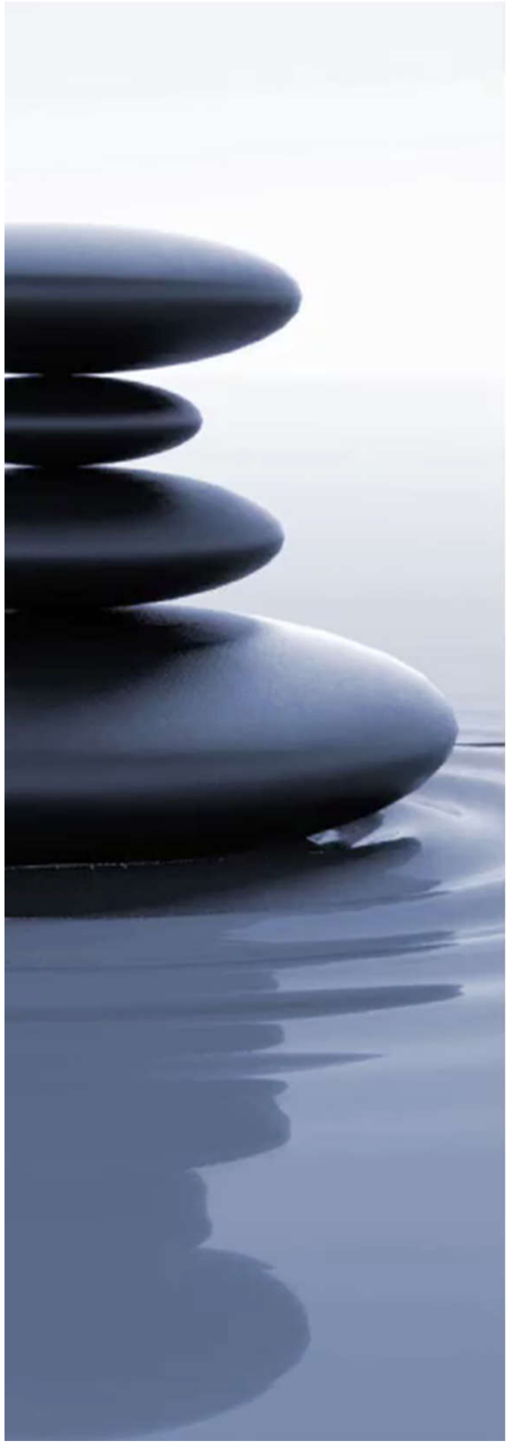
Image processing and computer vision, for face recognition, motion detection, and object detection

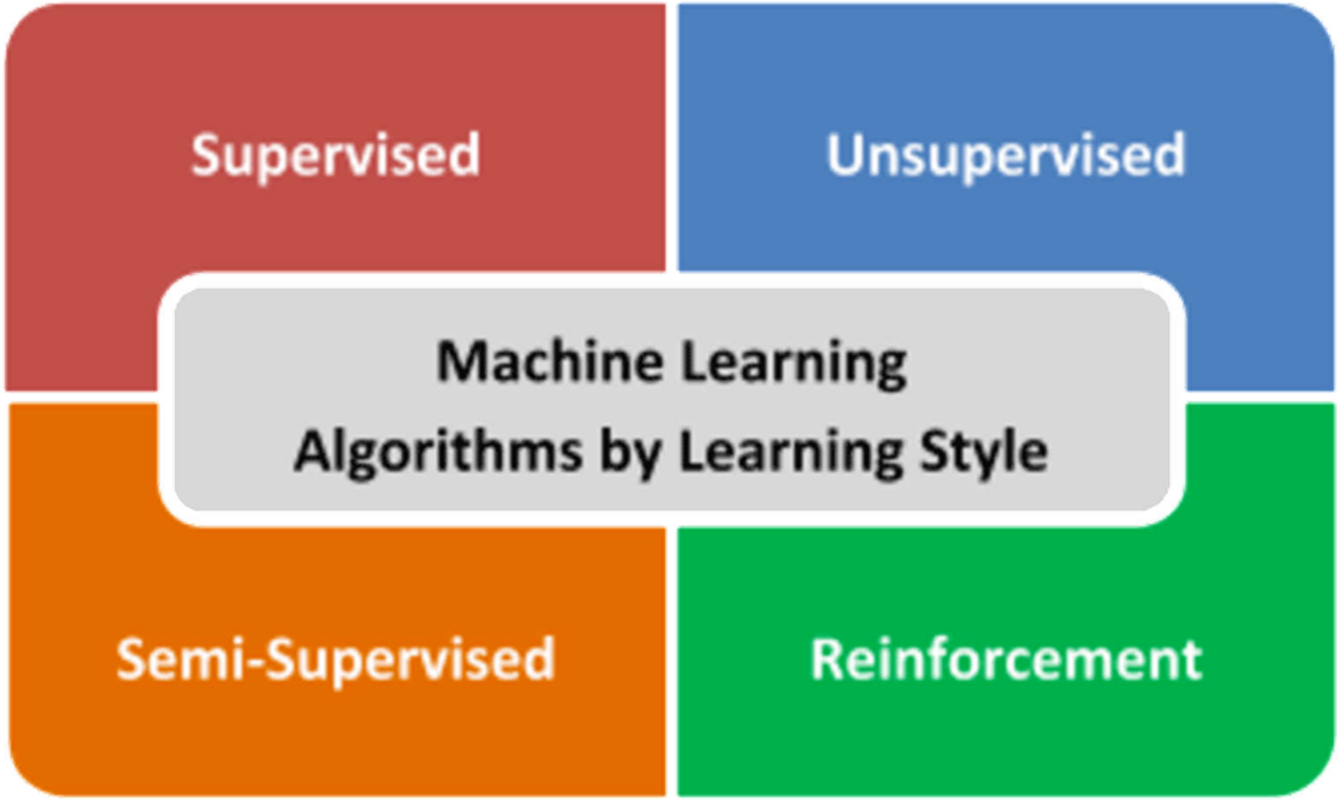
Computational biology, for tumor detection, drug discovery, and DNA sequencing

Energy production, for price and load forecasting

Automotive, aerospace, and manufacturing, for predictive maintenance

Natural language processing, for voice recognition applications





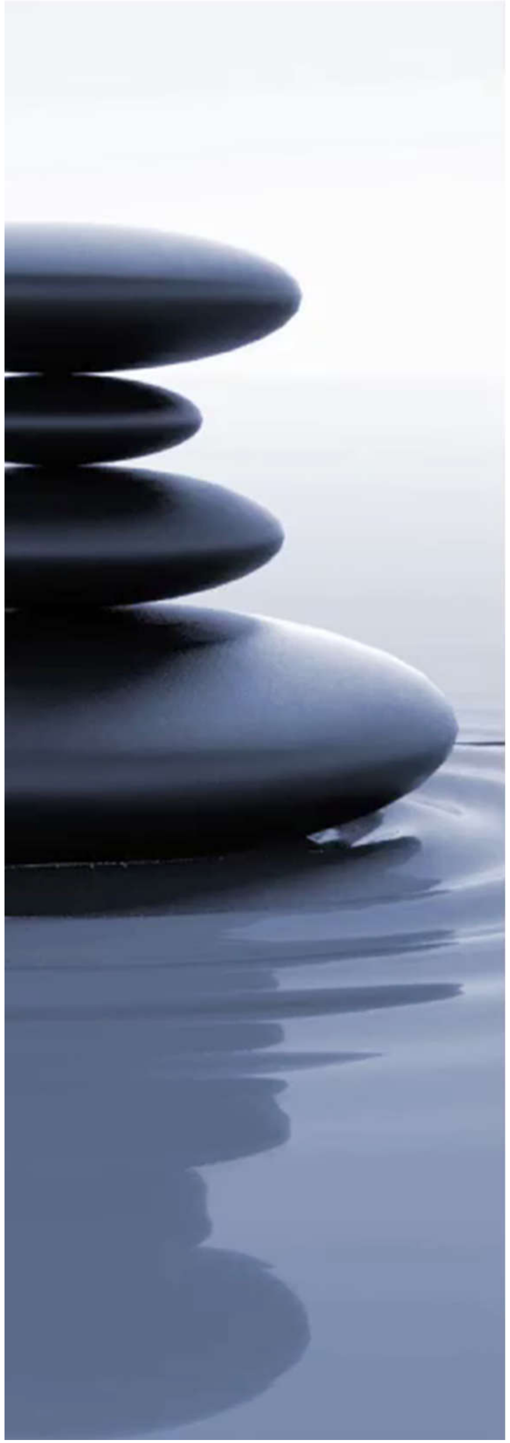
Supervised

Unsupervised

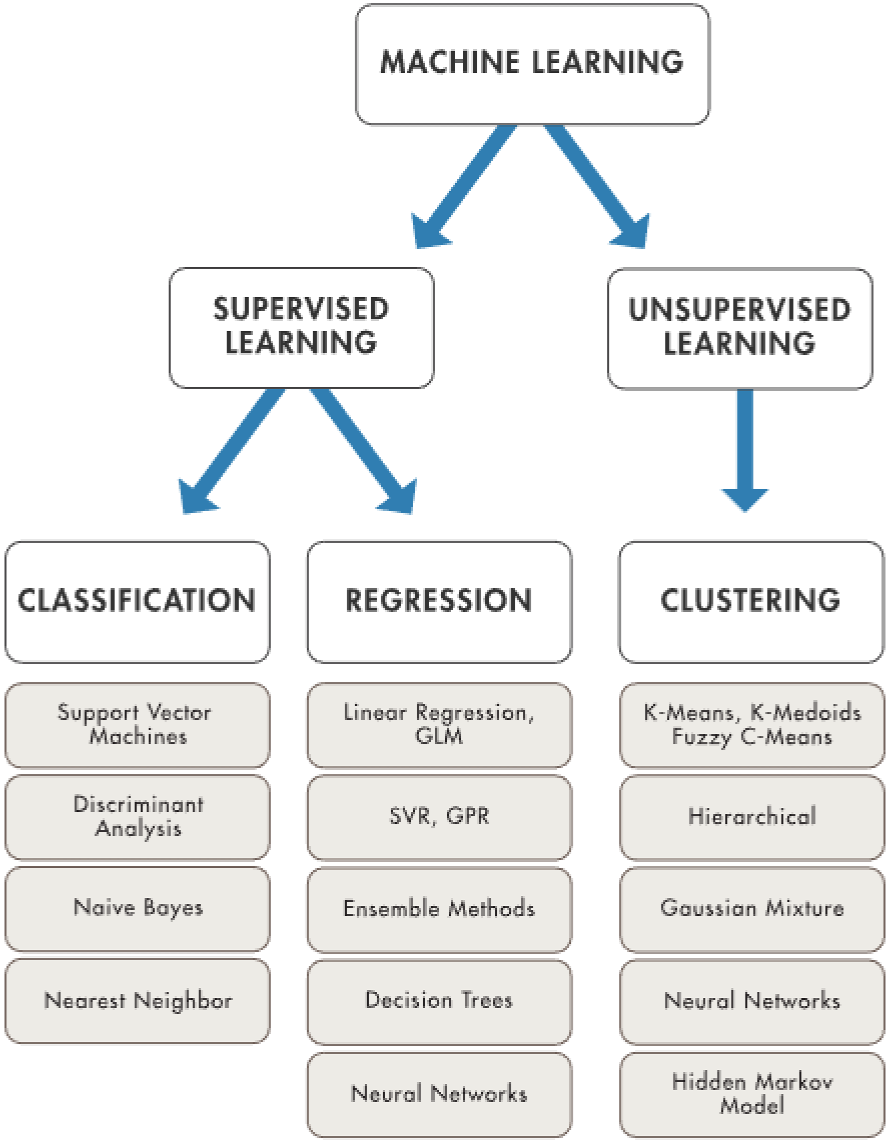
Machine Learning
Algorithms by Learning Style

Semi-Supervised

Reinforcement

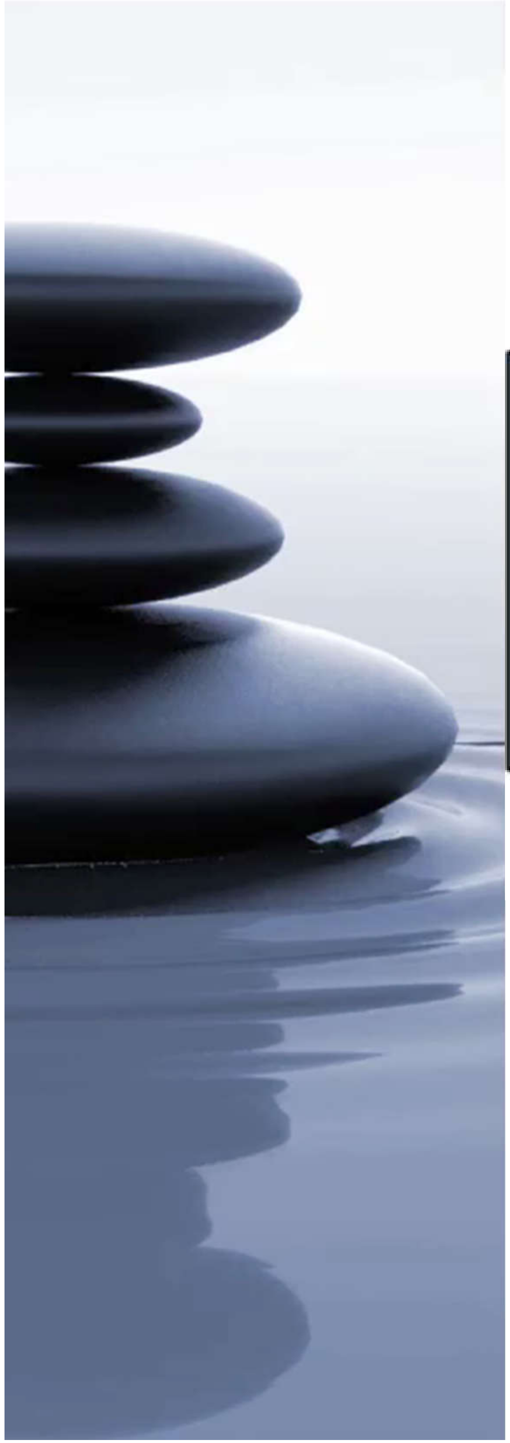


How Machine Learning Works

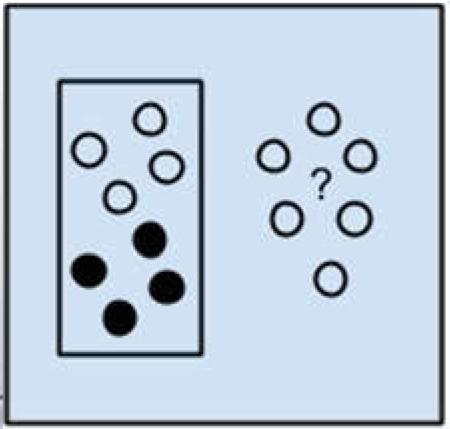


MACHINE LEARNING

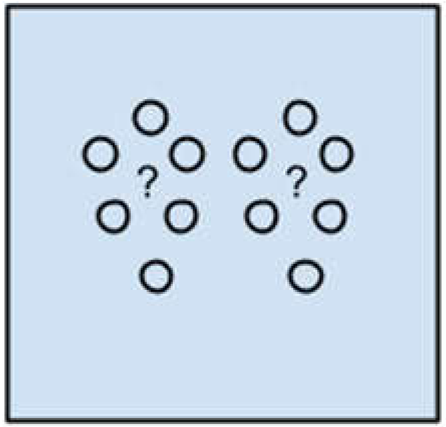
- SUPERVISED LEARNING
  - CLASSIFICATION
    - Support Vector Machines
    - Discriminant Analysis
    - Naive Bayes
    - Nearest Neighbor
  - REGRESSION
    - Linear Regression, GLM
    - SVR, GPR
    - Ensemble Methods
    - Decision Trees
    - Neural Networks
- UNSUPERVISED LEARNING
  - CLUSTERING
    - K-Means, K-Medoids Fuzzy C-Means
    - Hierarchical
    - Gaussian Mixture
    - Neural Networks
    - Hidden Markov Model



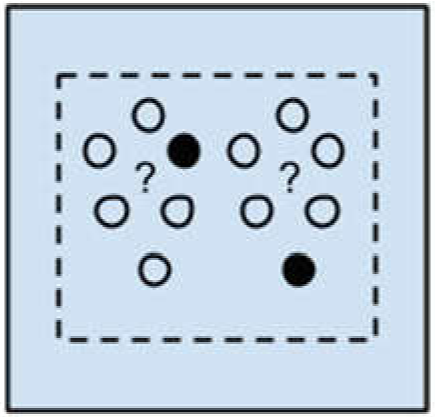
How Machine Learning Works



Supervised Learning Algorithms

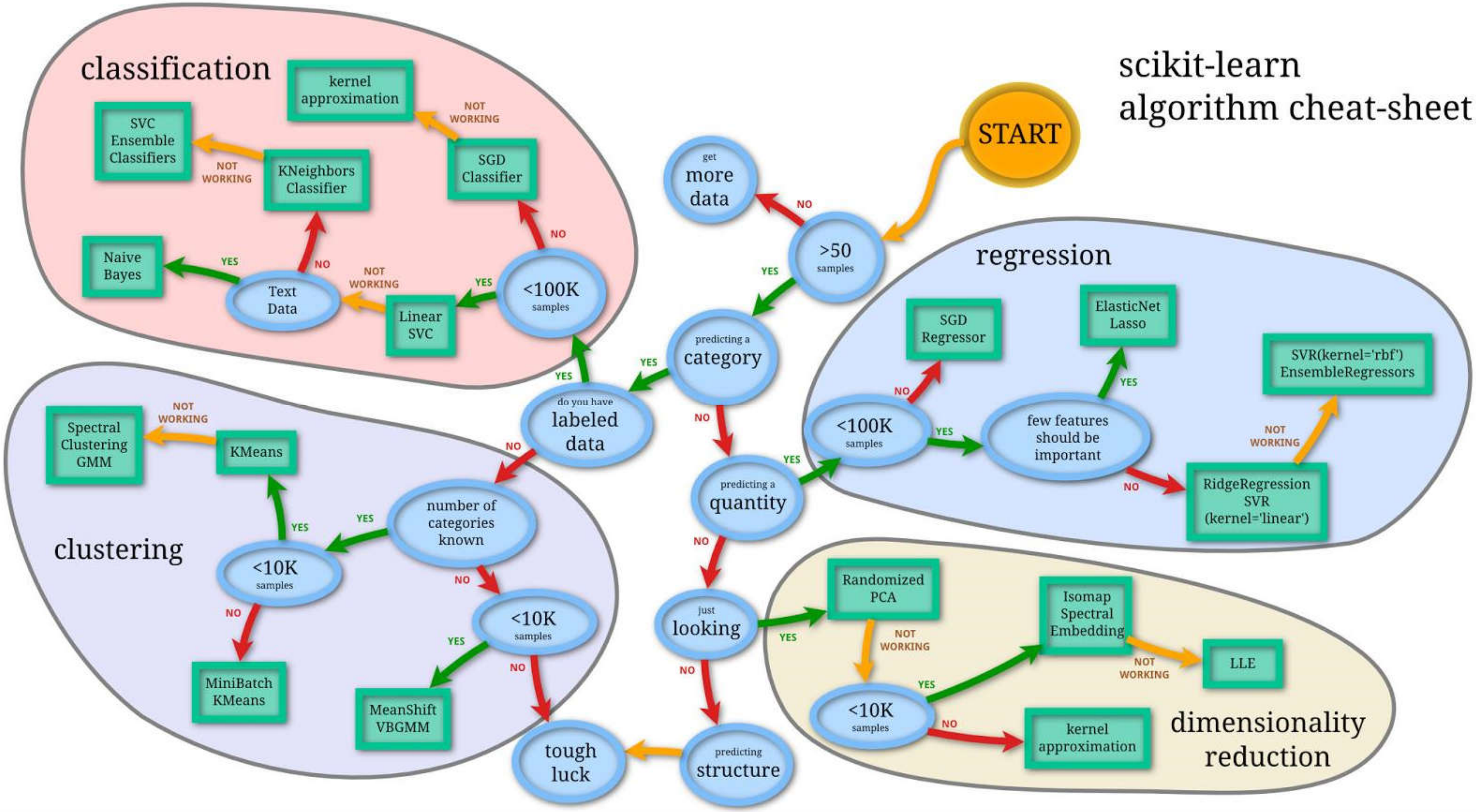


Unsupervised Learning Algorithms

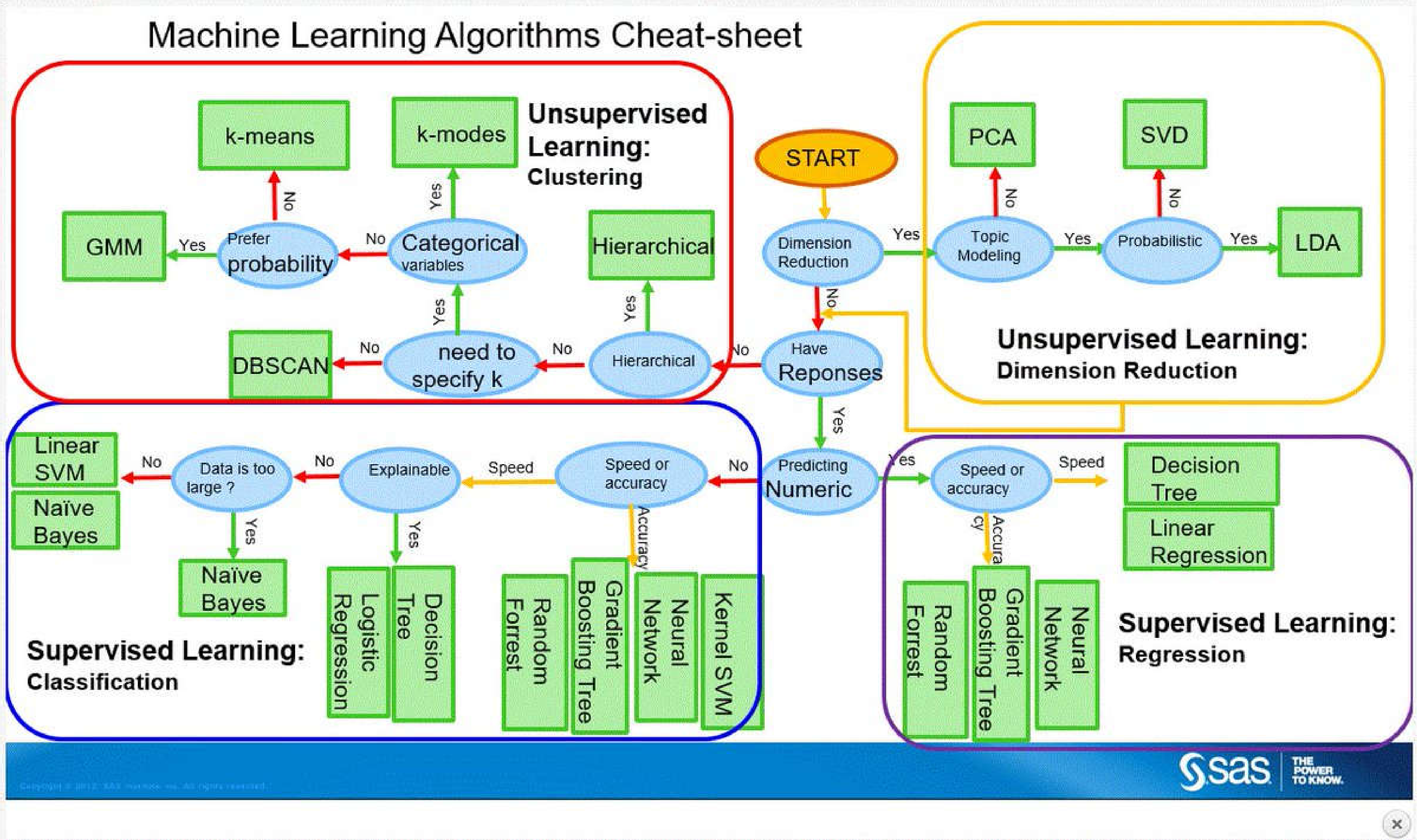


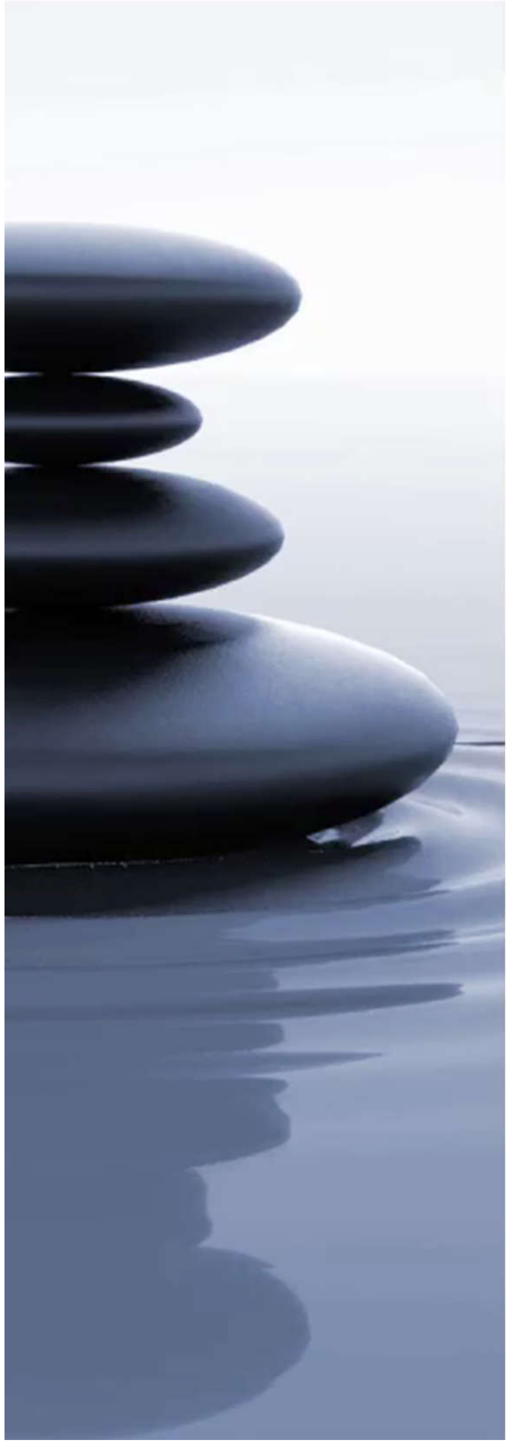
Semi-supervised Learning Algorithms

Machine Learning Styles



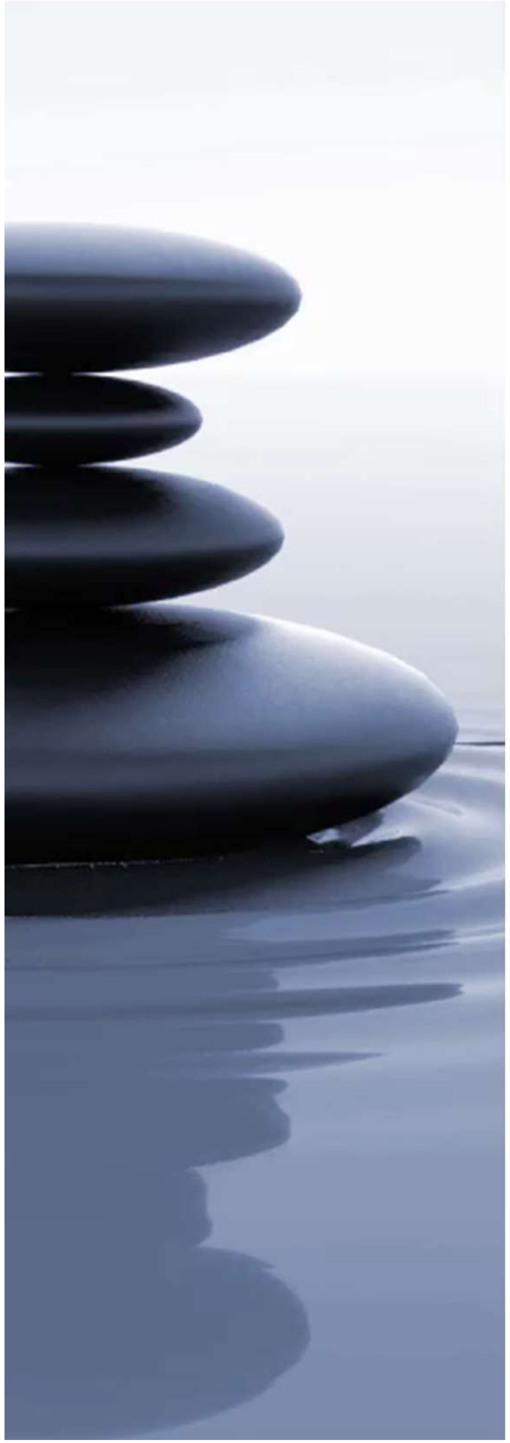
Machine Learning Styles


Machine Learning Algorithms Cheat-sheet

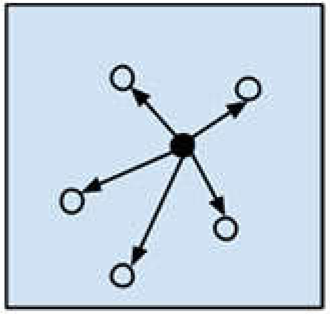


Machine Learning Algorithms by function

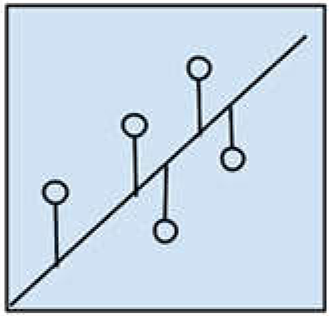
- Regression Algorithms
- Instance-based Algorithms
- Regularization Algorithms
- Decision Tree Algorithms
- Bayesian Algorithms
- Clustering Algorithms
- Association Rule Learning Algorithms
- Artificial Neural Network Algorithms
- Deep Learning Algorithms
- Dimensionality Reduction Algorithms
- Ensemble Algorithms



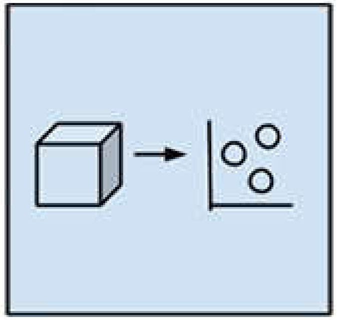
Machine Learning Algorithms by function



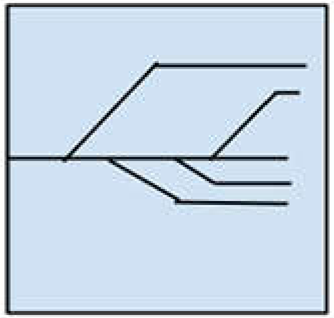
Instance-based Algorithms



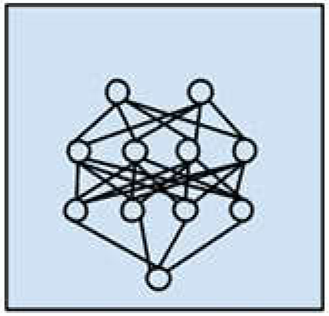
Regression Algorithms



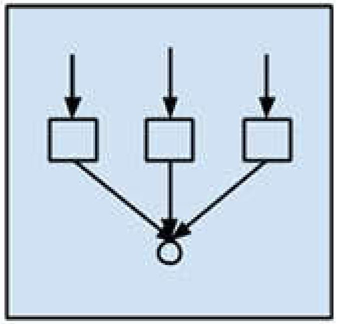
Dimensional Reduction Algorithms



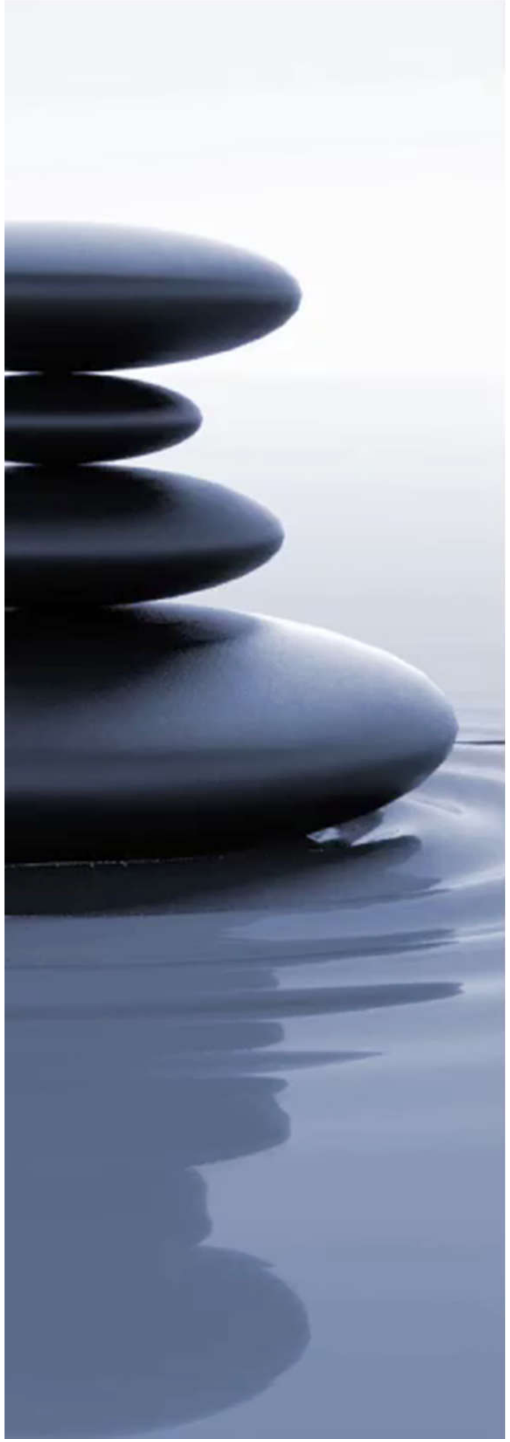
Regularization Algorithms



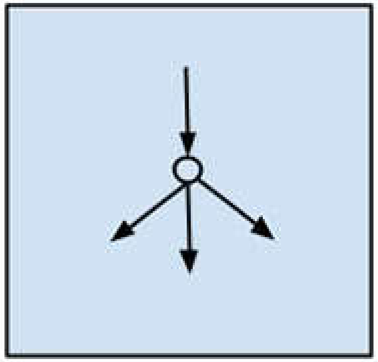
Deep Learning Algorithms



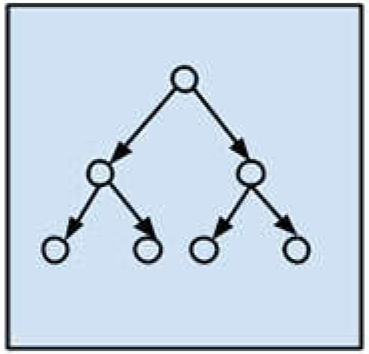
Ensemble Algorithms



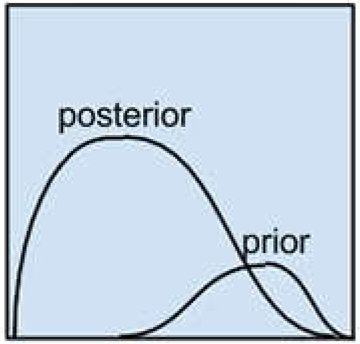
Machine Learning Algorithms by function



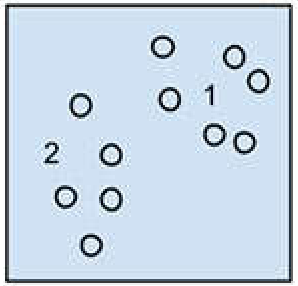
Artificial Neural Network Algorithms



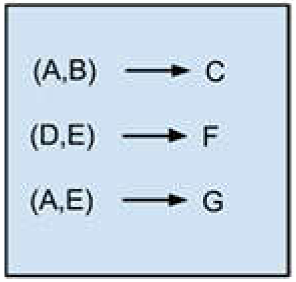
Decision Tree Algorithms



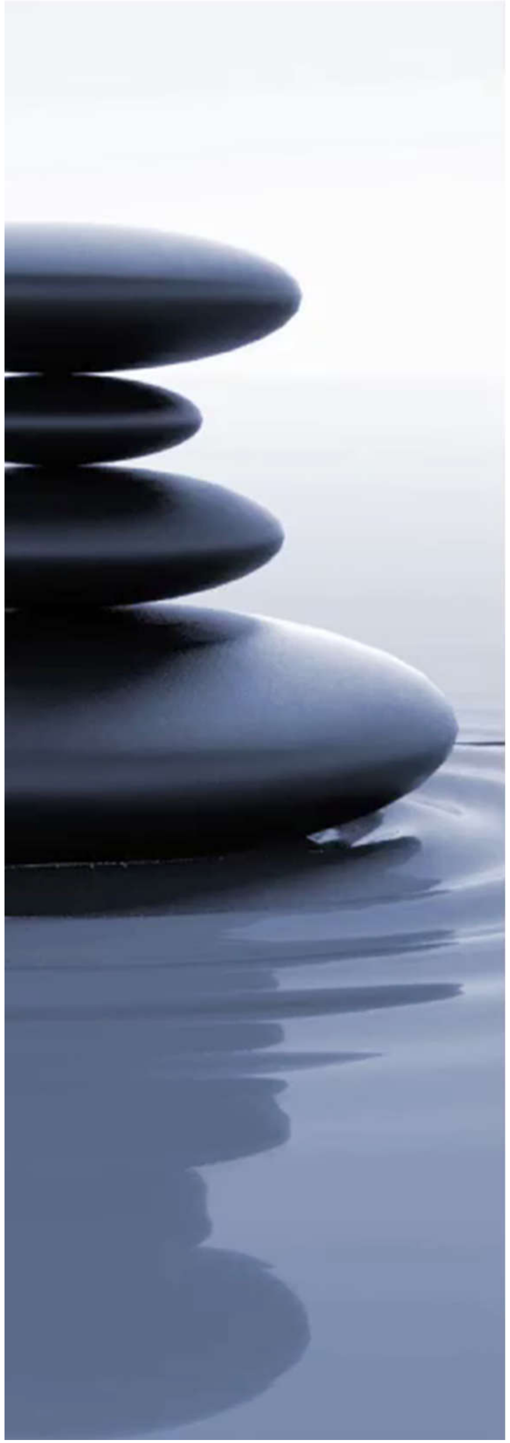
Bayesian Algorithms



Clustering Algorithms



Association Rule Learning Algorithms



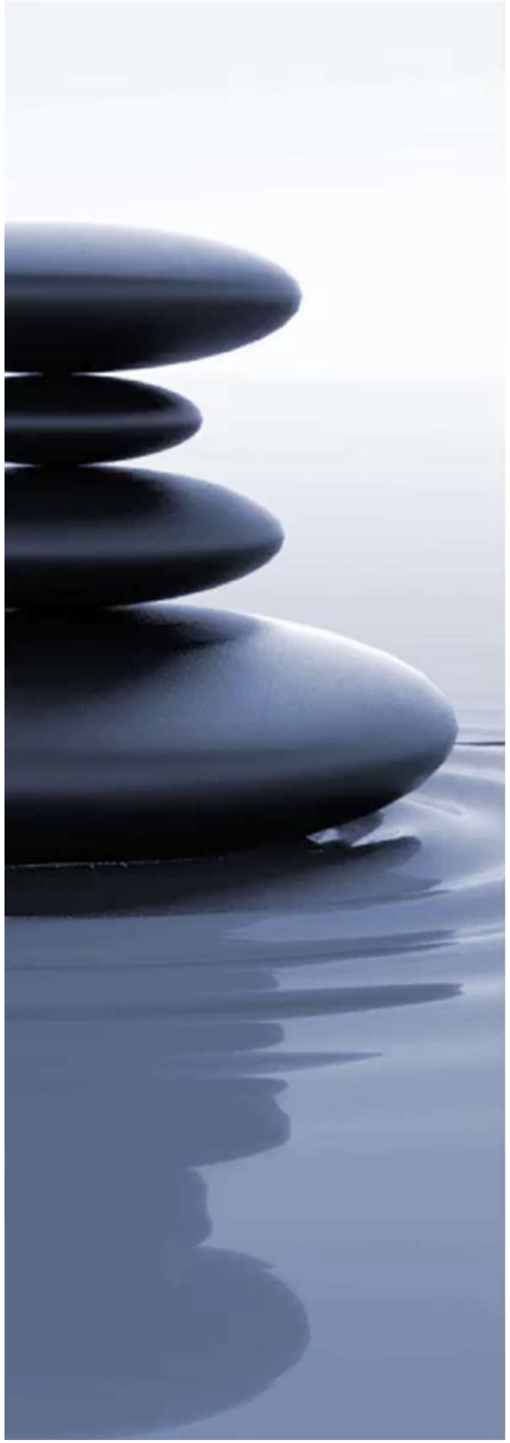
Basic concepts

Observation: input in each problem, denoted by **x**. Observation usually takes the form of a vector **x** = (x_1, x_2, ..., x_n) _ **feature vector**. Each x_i is a **feature**.

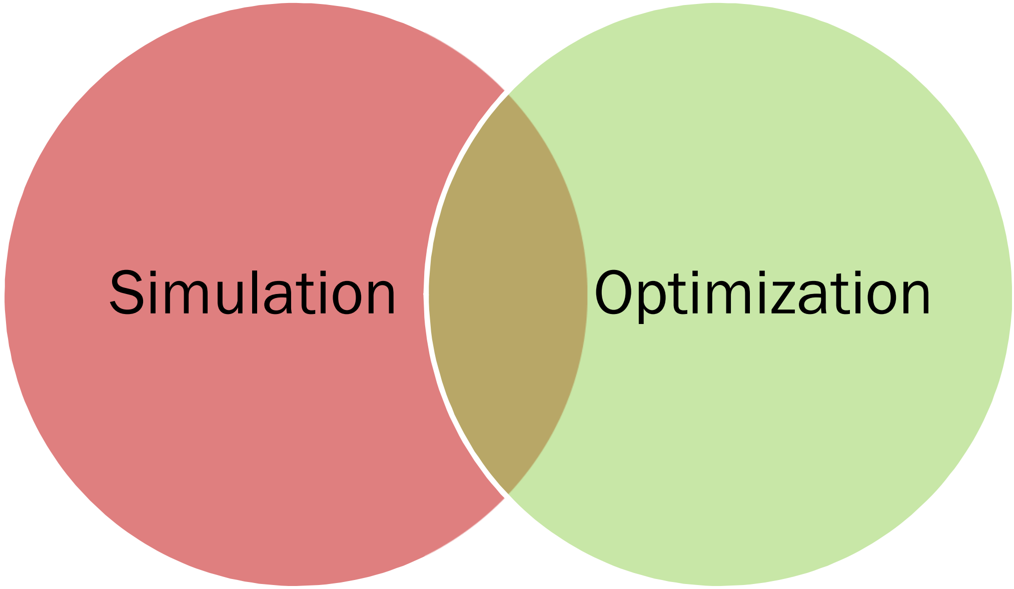
Label: output of each problem, denoted by *y*. Each observation will have a corresponding label.

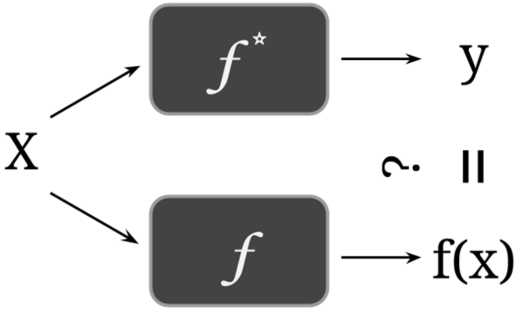
Model: Model is a function *f(*x*)*, take on an observation **x** and return a label *y=f(*x*)*.

Parameter: Everything of the model is used to calculate the output. People often gather all the parameters of a model into a vector, usually denoted by **w.** The notation $f_w(x)$ is used to refer to a model already defined parameters.

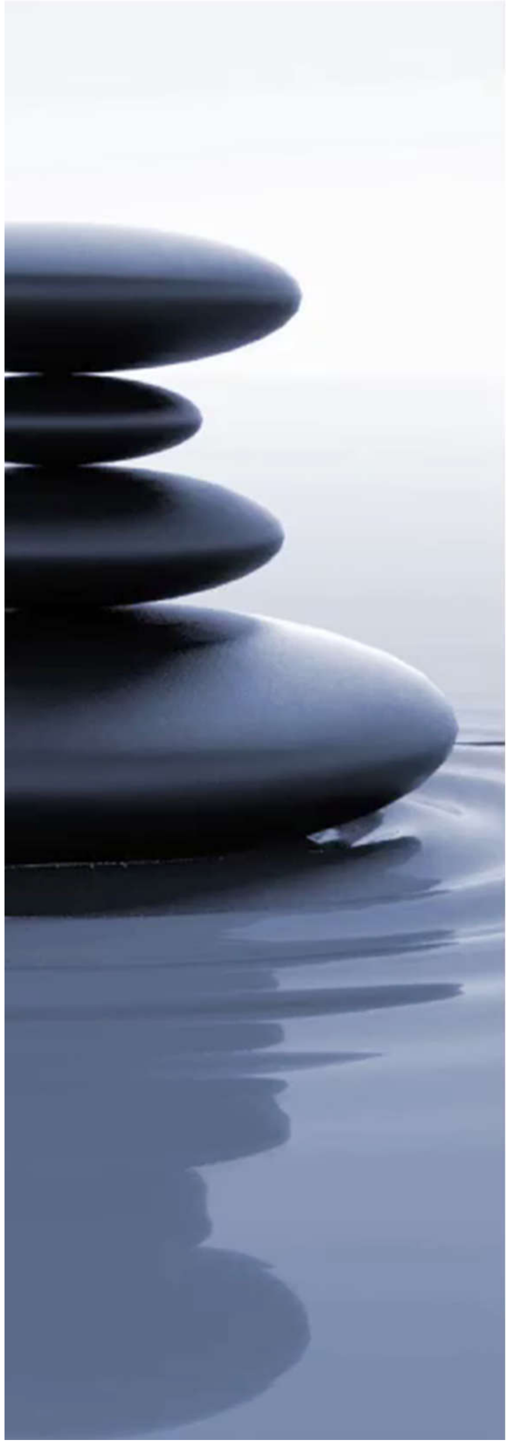


Perspectives of Machine Learning



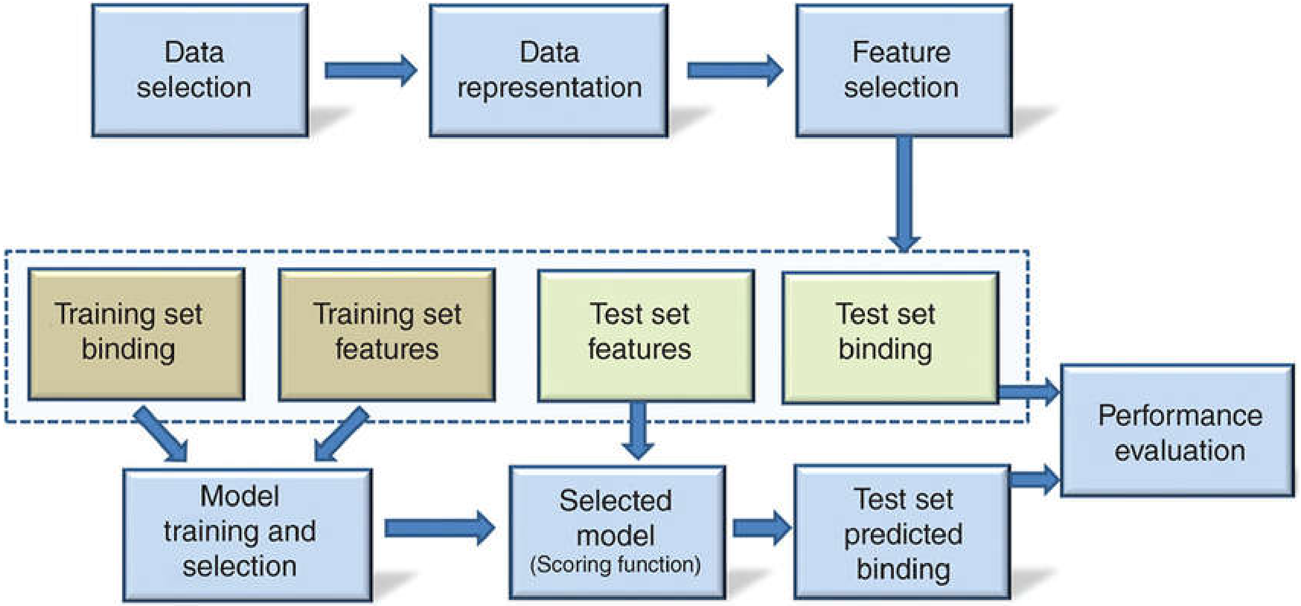


$$\mathcal{L}_{D_{test}}(f_w) = \frac{1}{|D_{test}|} \sum_{(x,y)\in D_{test}} L\left(f_w(x), y\right)$$



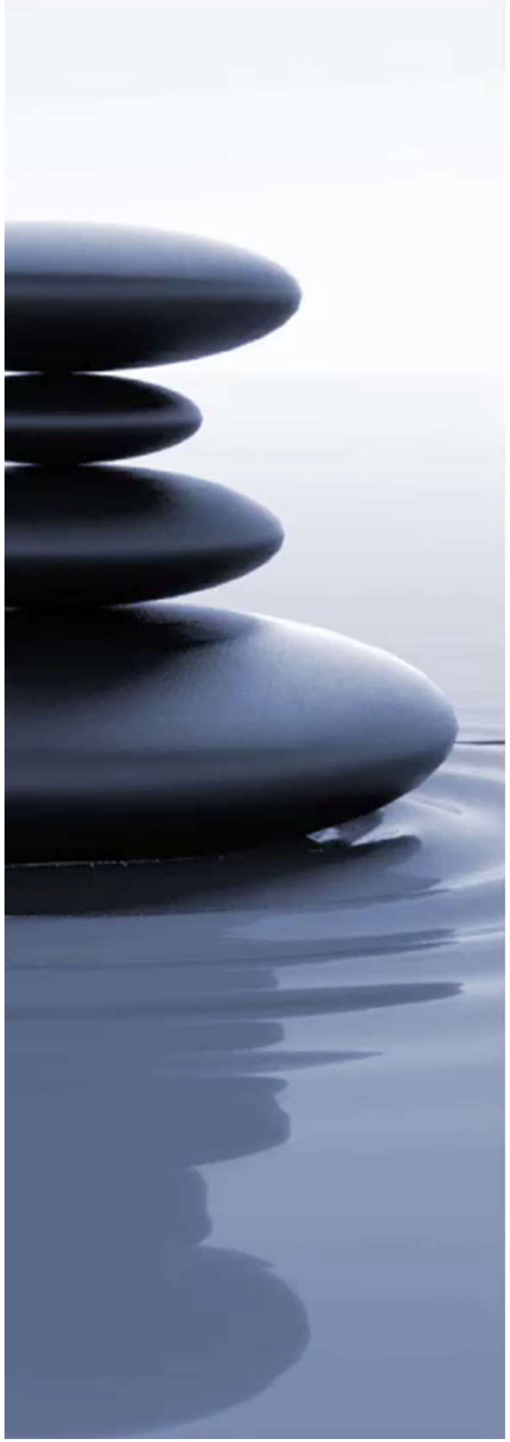
Objective Function

Machine learning data flow



Train : Find the model f_w to minimize the objective function on the training set. $\mathcal{L}_{D_{train}}(f_w)$

Test : Noticing the goodness of the model f_w, the value of the evaluation function on the test set. $\mathcal{L}_{D_{test}}(f_w)$



Objective Function

The form of objective functions is

objective=average_loss+regularizer

So,

$$\mathcal{L}_{D_{train}}(f_w) = \frac{1}{|D_{train}|} \sum_{(x,y)\in D_{train}} L(f_w(x), y) + \lambda R(f_w)$$

Where

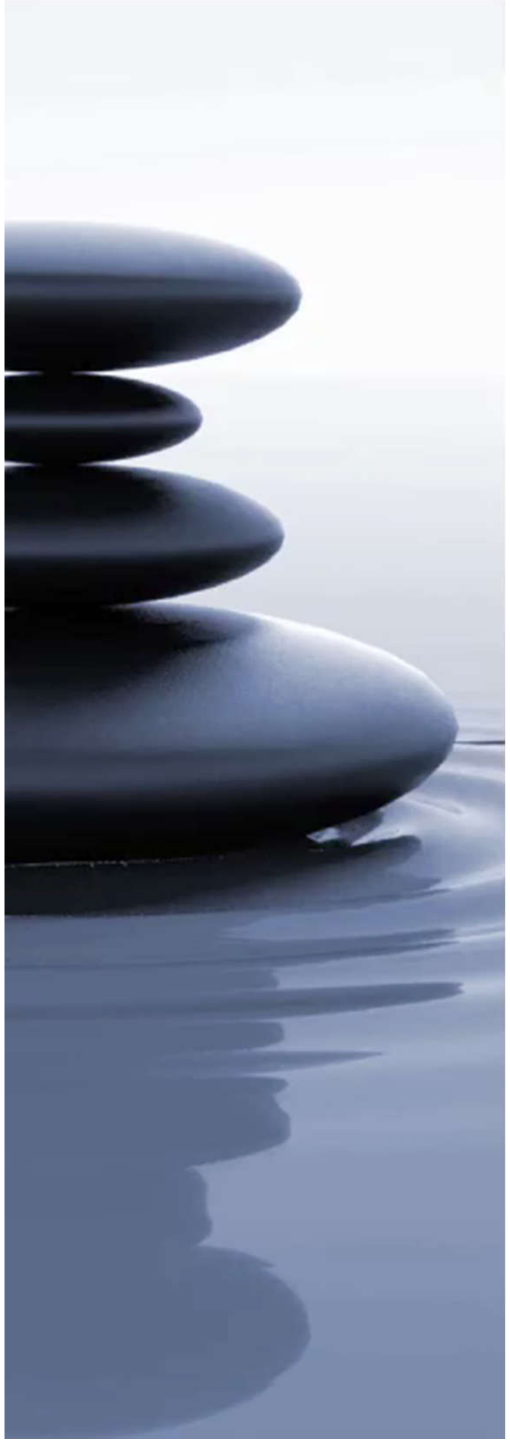
$\frac{1}{|D_{train}|} \sum_{(x,y)\in D_{train}} L(f_w(x), y)$ is an average loss function

$R(f_w)$ is a regularizer

λ is a regularization constant

The goal of training is to find a model that minimizes the objective function

$$w^* = \arg\min_w \mathcal{L}_{D_{train}}(w)$$



Noise

When we want to look for a model "entirely accurate predictions" on the training set, refers to the use of an objective function without regularizer:

$$\mathcal{L}^{ERM}_{D_{train}}(f_w) = \frac{1}{|D_{train}|} \sum_{(x,y)\in D_{train}} L(f_w(x), y)$$

This is called the empirical risk minimization (ERM)

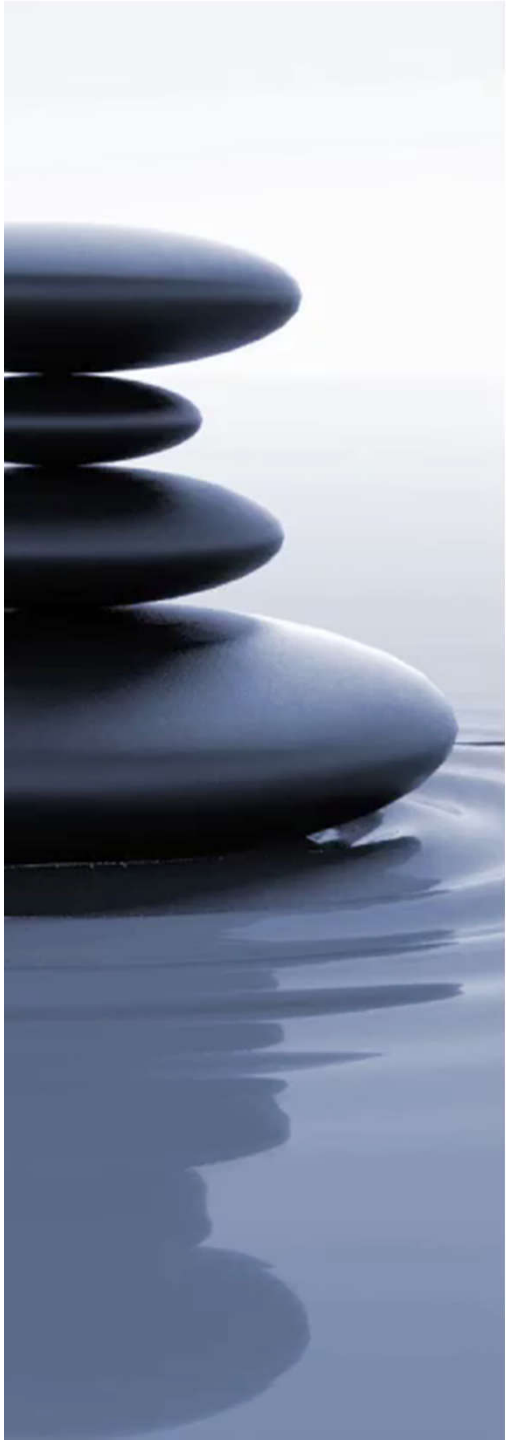
For any pair of data (x, y), one can describe its distortion process as follows:

$$\tilde{x} = x + \epsilon_x$$

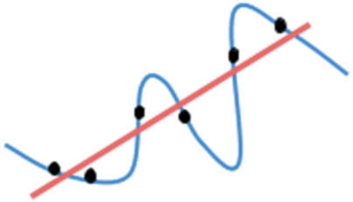
$$\tilde{y} = f(\tilde{x}) + \epsilon_y$$

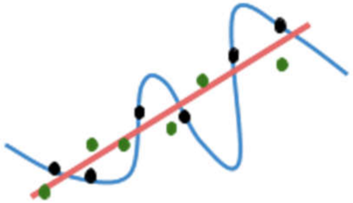
where ϵ_x and ϵ_y are noise of x and y respectively.

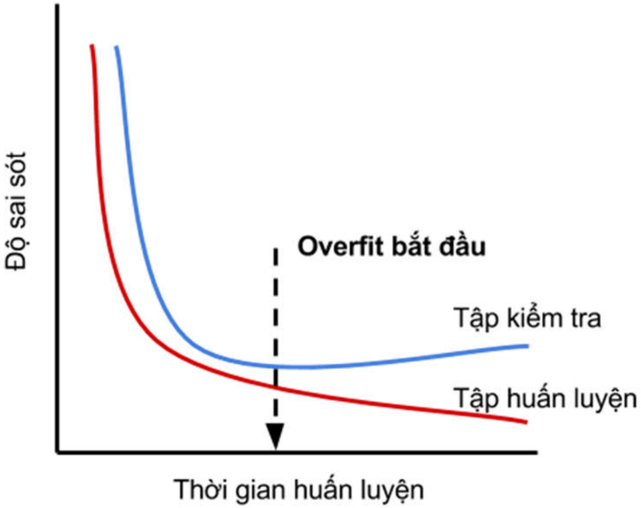
Finally, the actual data used for training and testing is ($\tilde{x}$, $\tilde{y}$), an error version of (x, y)



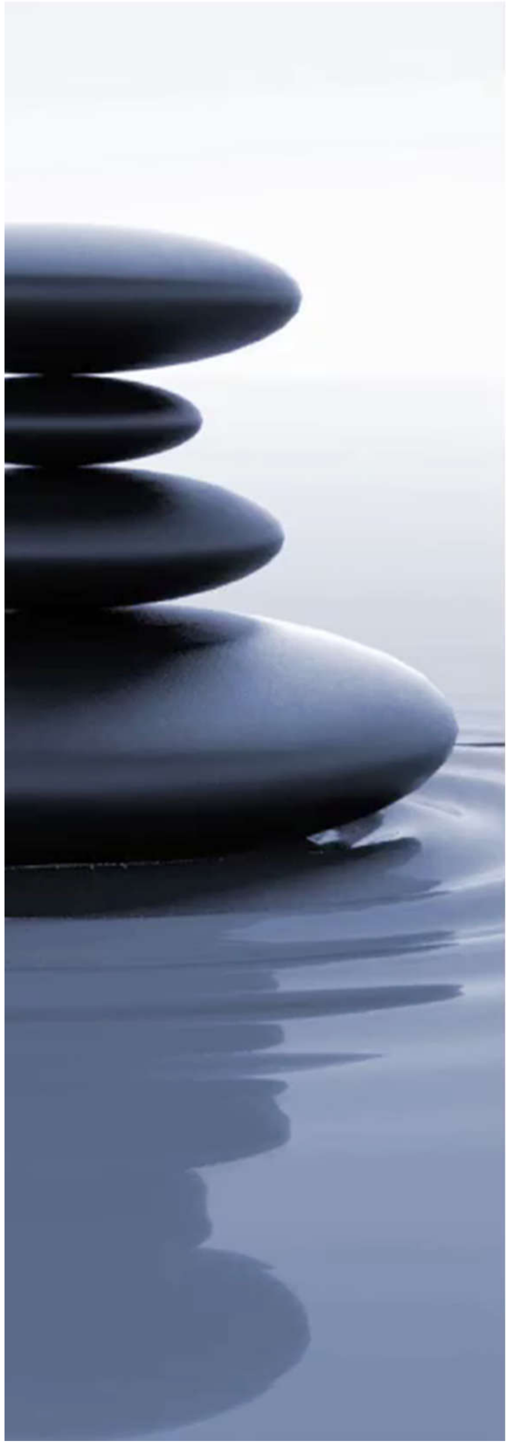
Overfitting







There are 2 techniques used to avoid overfitting: early stopping and regularizer.

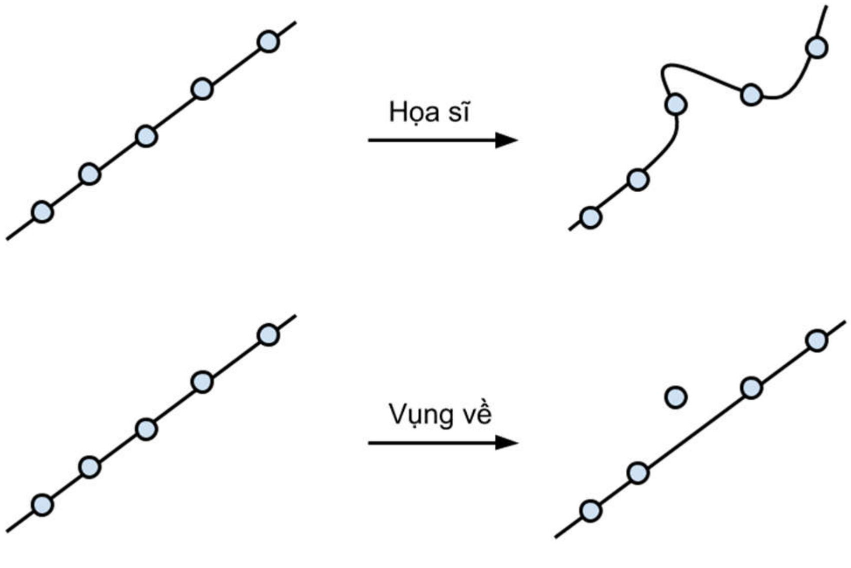


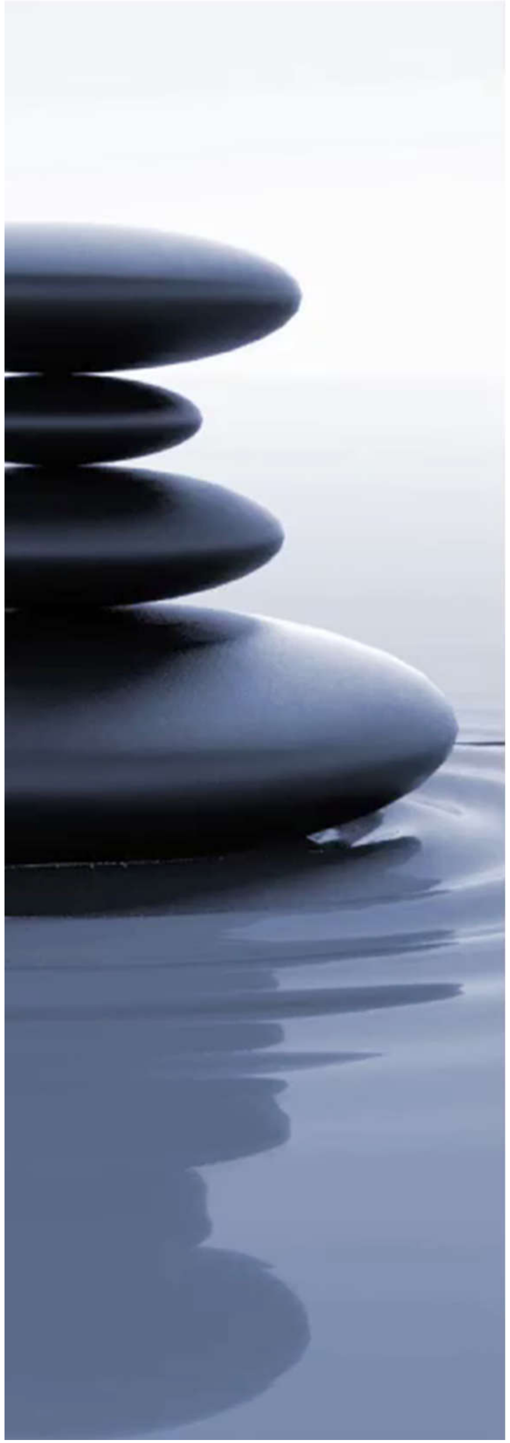
Causes of overfitting

Use ERM as an objective function

Limitations on data: The more training data, the less likely it is to overfitting.

Model too "strong“: a model too strong when it has the ability to simulate very complex relationships between observation and label





Regularized loss minimization

Main idea: Add to the ERM objective function a regularizer that limits the power of the model.

Approach: Assume that you have chosen a “too strong” model. There is no need to change the form of the model, it is possible to restrict its power by limiting parameter space (the space of parameters) of the model.

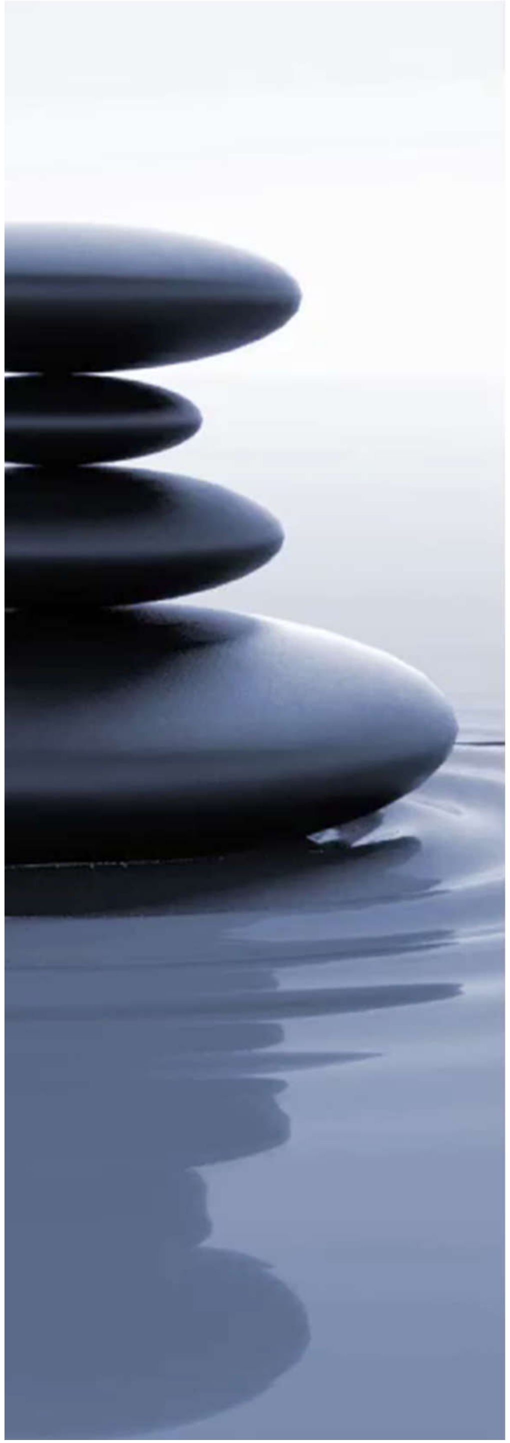
Technique: If the parameter w is a real vector with d dimensions, the set of these values that can be received, also called the parameter space of w, is the set of all vectors with d-dimensional real numbers, denoted by $\Re^d$. To limit this space, we need a mechanism to narrow the domain of each dimension.

$$\mathcal{L}_{D_{train}}(f_w) = \mathcal{L}_{D_{train}}^{ERM} + \lambda R(w)$$

R(w) most commonly is the norm of vectors.

1-norm (L1-norm): $R(w) = ||w||_1 = \sum_{i=1}^{d} |w_i|$

squared 2-norm (L2-norm): $R(w) = ||w||_2^2 = \sum_{i=1}^{d} w_i^2$

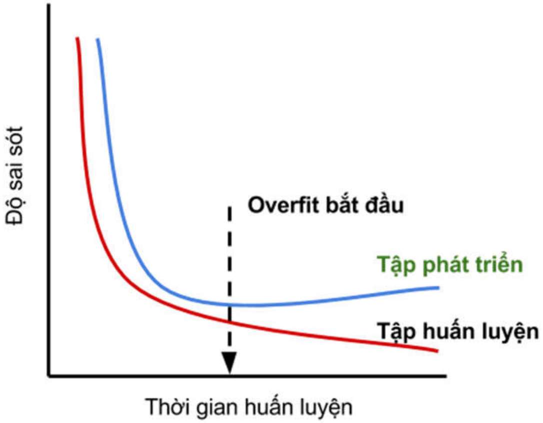


Refine hyperparameters

Training set

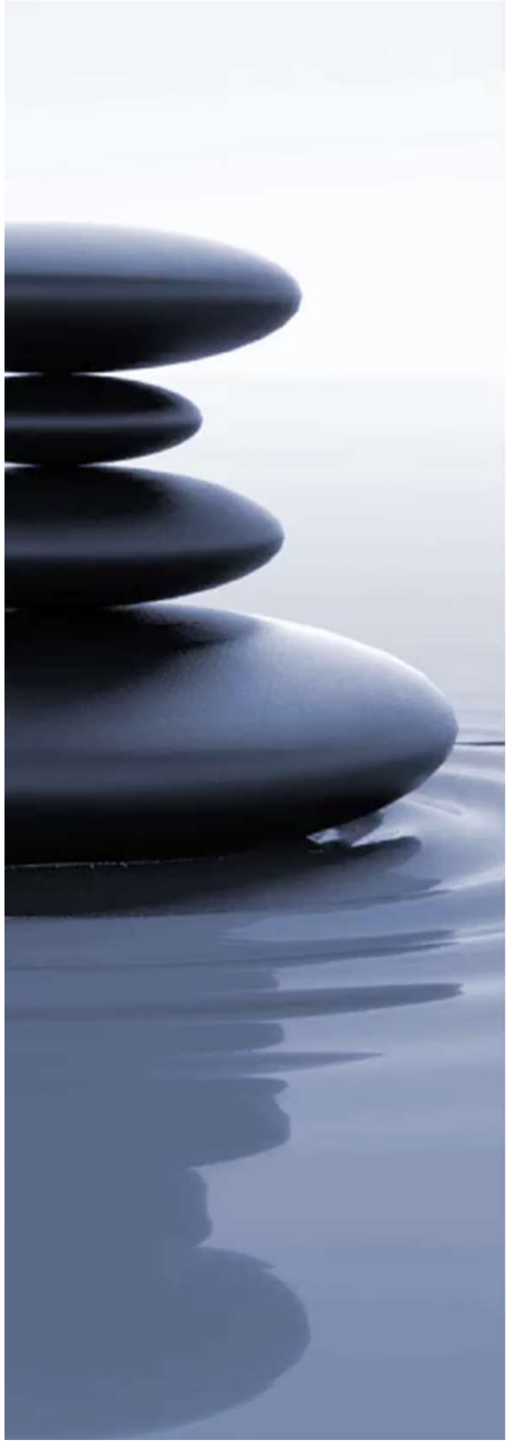
Test set

Development set or validation set



The ratio among train:dev:test is 8:1:1.

K-fold cross validation is used when there is not enough data to extract a good development set.



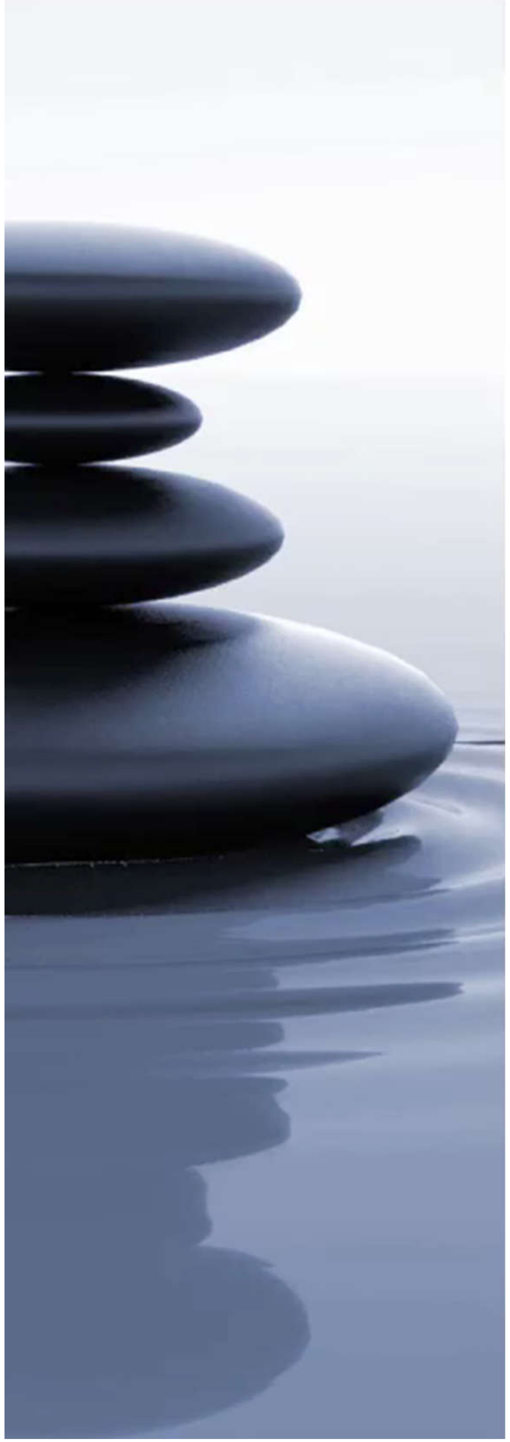
General machine learning algorithm

Step 1: Use the development set to refine the hyperparameter of the model.

- Training: Find w to minimize the objective function. Track the learning curve to apply early stopping ASAP.

- Evaluation: notice the value of the evaluation function on the development set.

Step 2: Evaluate on the test set. With the w* giving the best results on the development set, notice the evaluation function on the test set.



Loss function

Loss function is used to measure the degree of fit.

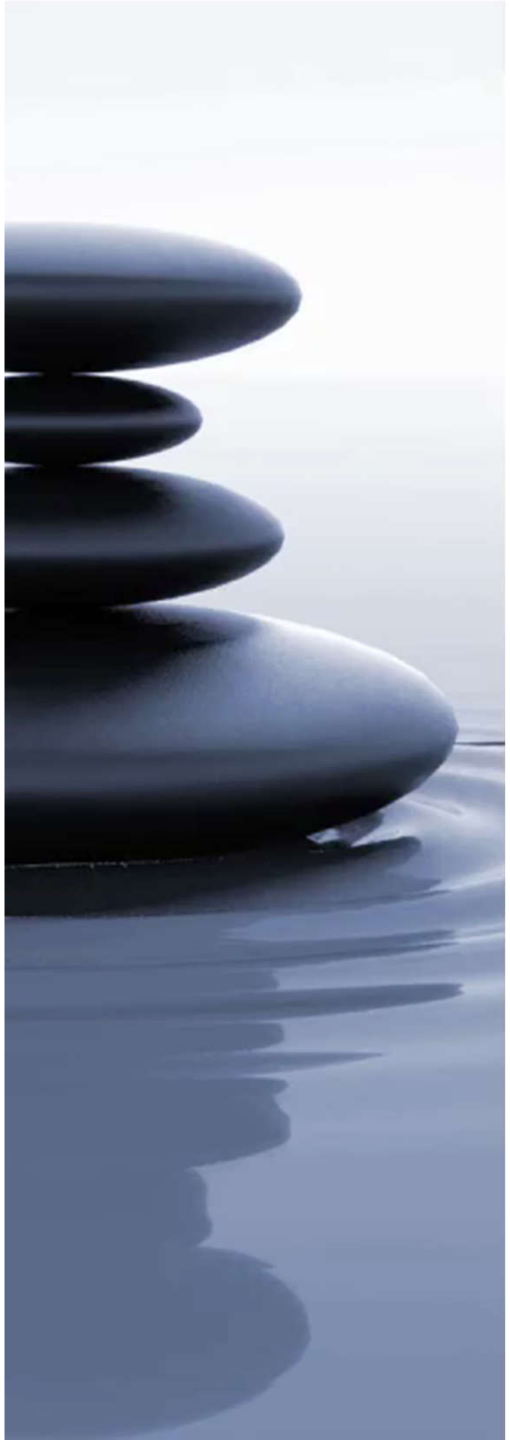
Loss function is denoted by L, which is the core of the evaluation function and the objective function. Specifically, in the common formula:

$$\mathcal{L}_D(f_w) = \frac{1}{|D|} \sum_{(x,y)\in D} L\left(f_w(x), y\right)$$

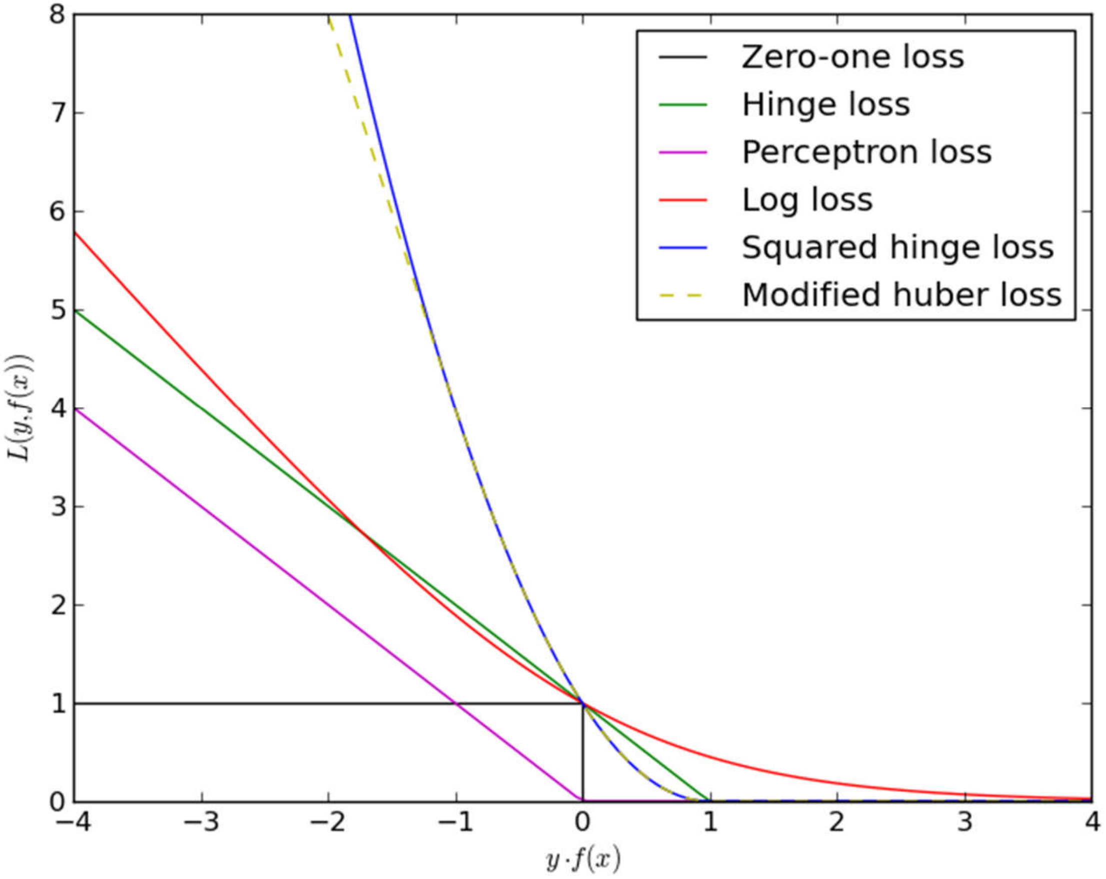
Loss function returns a nonnegative real number representing the difference between two quantities: $\tilde{y}$, label is predicted and y, label is correct.

$$L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$$

$$\nabla L = \tfrac{1}{2} \times 2 \times (\hat{y} - y) = \hat{y} - y.$$



Basic Loss functions for binary classification



Any Questions?

✉ hvusynh@hcmiu.edu.vn