

Outlines

Course Syllabus

What Is Data Science

Machine Learning

Mathematical Representation

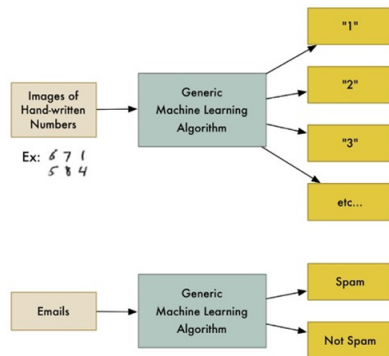
Conclusion

Definition

- Artificial Intelligence (AI) : learning from experiences (data), and improve the computer program adaptively
- Mathematics : Learning the underlying model from data, and generalize the model to adapt new data

We define *machine learning* as a set of methods that can automatically **detect patterns in data**, and then use the uncovered patterns to **predict future data**, or to **perform other kinds of decision making under uncertainty** (such as planning how to collect more data!).

— 《Machine Learning: A probabilistic perspective》



Related Areas

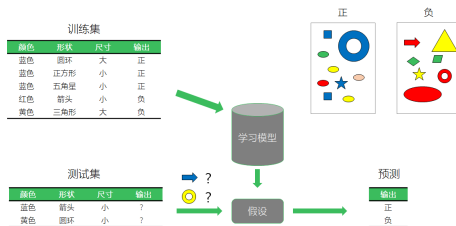
- Control theory : optimize the cost with optimal control parameters
- Information theory : entropy, optimal coding with best information
- Psychology : reference for machine learning algorithms
- Neuroscience : artificial neural network
- Biology : genetic algorithms
- Theory of Computing : study the computational complexity
- Statistics : large-sample limiting behavior, statistical learning theory
- Artificial Intelligence : symbolic computing
- Bayesian theory : conditionally probabilistic network

Supervised and Unsupervised Learning

- Supervised learning : classification, regression
- Unsupervised learning : density estimation, clustering, dimensionality reduction
- Semi-supervised learning : with missing data, e.g., EM ; self-supervised learning, learn the missing part of images, inpainting
- Reinforcement learning : play games, e.g., Go, StarCraft ; robotics ; auto-steering

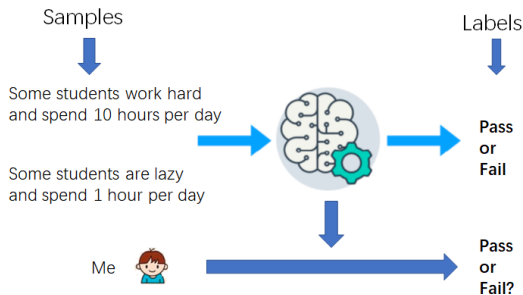
Supervised Learning

- Given labels of data : the labels could be symbols (spam or non-spam), integers (0 or 1), real numbers, etc.
- Training : find the optimal parameters (or model) to minimize the error between the prediction and target
- Classification : SVM, KNN, Decision tree, etc.
- Regression : linear regression, CART, etc.



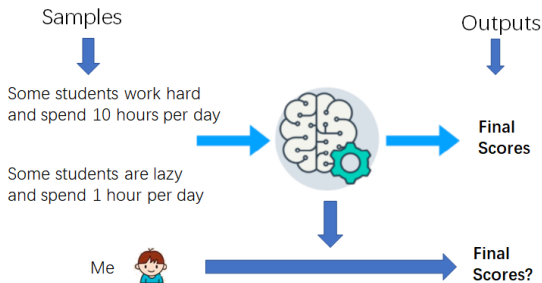
Classification

- Output is discrete
- Examples : given the study hours, in-class performance, and final grades (Pass or Fail) of past students, can you predict the final grades of the current students based on their study hours and in-class performance ?
- Applications : Credit risk evaluation, clinical prediction of tumor, classification of protein functions, etc.



Regression

- Output is continuous
- Examples : given the study hours, in-class performance, and final scores of past students, can you predict the final scores of the current students based on their study hours and in-class performance?
- Applications : epidemiology, finance, investment analysis, etc.



Unsupervised Learning

- No labels
- Optimize the parameters based on some natural rules, e.g., cohesion or divergence
- Clustering : K-Means, SOM

