Applied Machine learning

Week 01

Introduction to ML

Course outline

1. Introduction + (Math & Stats Review)
2. Linear Regression (Single variable)
3. Linear Regression (Multiple variables) + Polynomial
4. Naive Bayes Classification
5. Logistic Regression
6. Regularization
7. Model Evaluation

8.~9. K-NN, Tree
10. Neural Network
11. Support Vector Machine (SVM)

12~13. Clustering (K-means, DBScan, ...)

14. Dimension Reduction, AutoEncoder

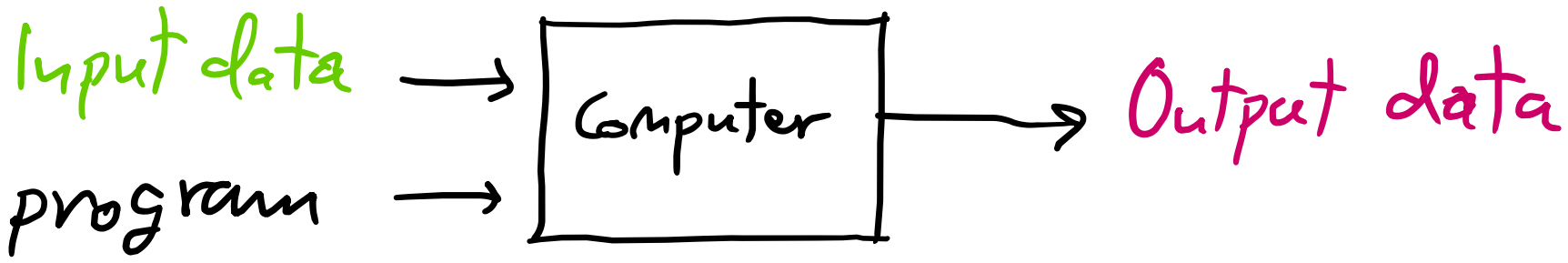
15. Interesting Topics

Week 1 outline

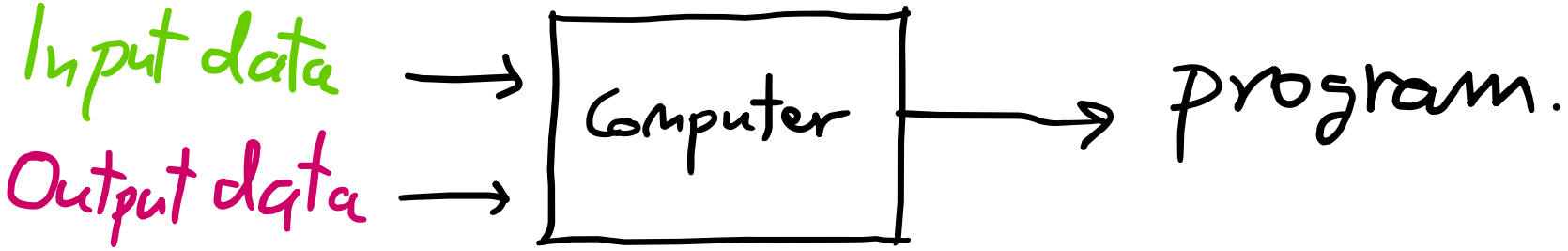
1. Introduction to Machine learning
2. Types of Machine learning
   - 2.1 Supervised learning
   - 2.2 Unsupervised learning
   - 2.3 Reinforcement learning
3. Machine learning process.

Machine learning Definition (1)

Traditional approach



Machine learning approach



Machine learning Definition (2)

- Arthur Samuel (1959)
  - ML → A Field of study that gives computers the ability to learn without explicitly programmed.

This Photo by Unknown Author is licensed under CC BY-SA

Machine learning Definition (3)

- Tom Mitchell (1998) Well-posed learning problem:

"A computer program is said to learn from Experience E with respect to some Task T and some Performance measure P,
If its performance on T, as measured by P, improves with experience E."

Machine learning Definition

- Task T: classifying handwritten digits from images
- Training experience E: dataset of digits given classifications, e.g., MNIST[4]
- Performance measure P : percentage of digits classified correctly

Machine learning Definition

T: Playing checkers
P: Percentage of games won against an arbitrary opponent
E: Playing practice games against itself

T: Driving on four-lane highways using vision sensors
P: Average distance traveled before a human-judged error
E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.
P: Percentage of email messages correctly classified.
E: Database of emails, some with human-given labels

2. Types of Machine learning

2.1 Supervise learning

- Given $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, $x_i \in R^d$ labeled data

  if $y_i \in R \rightarrow$ Regression problem.

  if $y_i \in \{C_1, C_2, \dots, C_K\} \rightarrow$ Classification problem. where C_K is a class K.

- Objective
  - Find a function f to predict x_i' in case of regression and to classify x_i' in case of classification.

2. Types of Machine learning

2.1 Supervise learning

- Algorithms
- Regression (E.g.)
  - Linear Regression
  - Polynomial Regression
- Classification (E.g.)
  - Naive Bayes
  - Decision Tree
  - SVM
  - Neural Network
  - ...

2. Types of Machine learning

2.2 Unsupervise learning

without labeled data.

- Given $D = \{ x_1, x_2, \ldots, x_N \}$, $x_i \in R^d$

- Objective - To understand and extracting pattern, structures, or relationships from unlabeled data.

2. Types of Machine learning

2.2 Unsupervise learning

- Primary goals
  - Clustering
  - Dimensionality Reduction
  - Anomaly Detection
  - Association Rule

2. Types of Machine learning

2.2 Unsupervise learning

- Algorithms
  - K-means
  - DBSCAN
  - PCA
  - ...

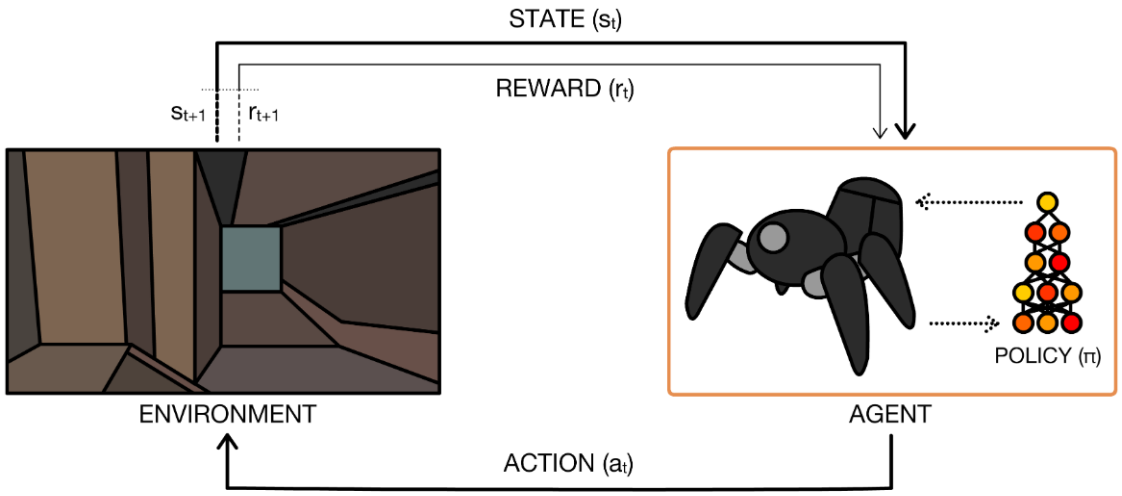
2. Types of Machine learning

2.3 Reinforcement learning

- Given a sequence of states, actions, and rewards
- Output an optimal policy (A mapping from states to actions)

Examples of applications

- Robotics
- Game playing



This Photo by Unknown Author is licensed under CC BY-SA

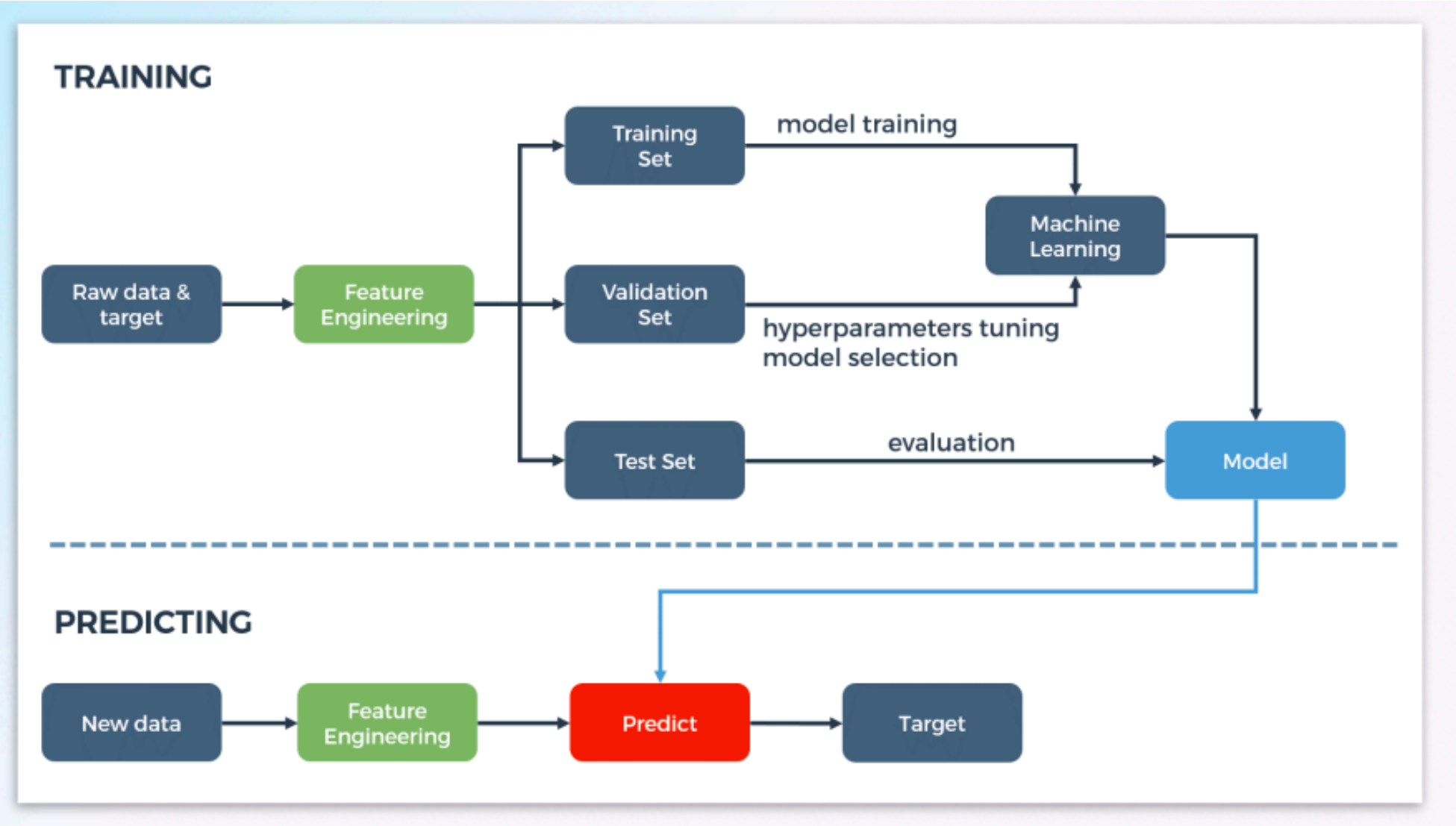
3. Machine learning Applications

- Email spam detection
- Face detection and recognition
- Sport analytics
- Zip codes recognition
- Credit card fraud detection
- Stock prediction
- Smart assistants (ChatGPT)
- Recommendations
- Self-driving cars
- ...

Machine Learning Process

1. What is the desired outcome? (Business Requirement)
2. What could the dataset look like? (E)
3. Is this a supervised/unsupervised/reinforcement problem? (T)
4. What algorithms would you use? (Solutions)
5. How would you measure success? (P)
6. How to maintain the generated models?
7. What are potential challenges or pitfalls?

Machine Learning Process



Components of Machine Learning Algorithms

1. Representation
2. Optimization
3. Evaluation

Components of Machine Learning Algorithms

1. Representations

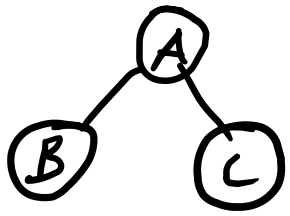
1.1 Numerical functions

- Linear Regression $y = \theta_0 + \theta_1 x$ / $\theta^T x$
- Logistic Regression $y = \sigma(\theta_0 + \theta_1 x)$ / $\sigma(\theta^T x)$

$\sigma \rightarrow$ sigmoid function

1.2 Symbolic functions

- Decision Tree



- Rule base

if(A==B)then C

Components of Machine Learning Algorithms

1. Representations

1.3 Instance-based functions

- k-Nearest Neighbors

1.4 Probabilistic Graphical Models

- Naive bayes

Components of Machine Learning Algorithms

2. <u>optimization methods</u>

- Gradient Descent
- Stochastic Gradient Descent
- RMSProp
- AdaGrad
- Newton's method
- Hessian Free method
- Conjugate Gradient
- ...

Components of Machine Learning Algorithms

3. <u>Evaluation</u> (Error metrics)

- Accuracy
- MSE
- MAE
- RMSE
- Precision, Recall, F1-Score
- Kappa
- MCC (Matthews Correlation Coefficient)

END