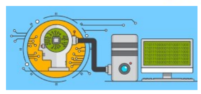
Applied Data Science

Session 15-28: Applied Machine Learning

Dr. Soharab Hossain Shaikh



1

Machine Learning = Learning from Data

- A computer program is said to **learn from experience E** with respect to some class of **tasks T** and **performance measure P** if its **performance at tasks in T**, as measured by P, **improves with experience E**.

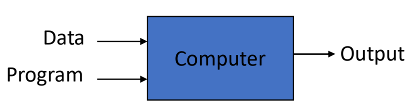
  –Tom Mitchell, Book on Machine Learning

- **Not just memorizing** : One aspect of machine learning is its **ability to modify itself when exposed to more data**; i.e. machine learning is **dynamic/adaptive** and **does not require human intervention** to make certain changes.
- Less brittle – more flexible
- Less reliant on human experts – less biased.

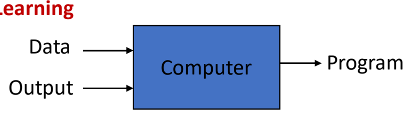
2

2

Traditional Programming



Machine Learning



5

3

What is Machine Learning?

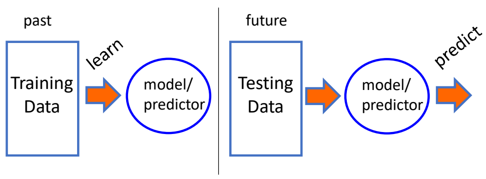
- Subset of Artificial Intelligence(AI)
- Gives "computers the **ability to learn without being explicitly programmed**" (Arthur Samuel, 1959)
- Focuses on development of computer programs that can **access data** and use it to **learn for themselves**.
- The process of learning begins with data in order to **find patterns in the data** and **make better decisions** in future (predictions) based on examples provided in data.
- Aim is to allow computers to learn automatically **without human intervention** and adjust actions accordingly.

4

4

Machine Learning is…

Machine learning is about predicting the future based on the past.



7

5

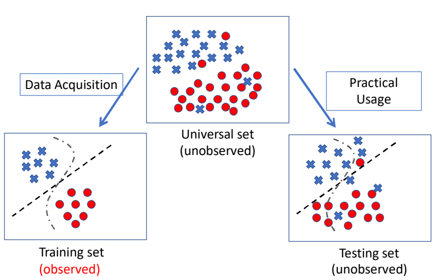
Machine Learning

- **Learning from Data.**
- The "learning" part of machine learning means that ML algorithms attempt to **optimize** along a certain dimension; i.e. they usually try to minimize error or maximize the likelihood of their predictions being true.
- Optimizing an **error/loss/cost function**.
- Learning a **mathematical equation representing the relationship between the inputs and output**.

6

6

Training and Testing

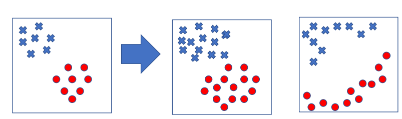


12

7

Training and Testing

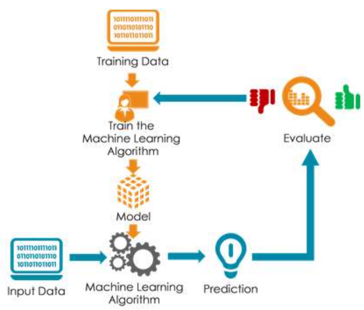
- **Training** is the process of **making the system able to learn**.
- No free lunch rule:
  - Training set and Testing set come from the **same distribution**
  - Need to make **some assumptions or bias**



13

8

Train, Evaluate, Predict



9

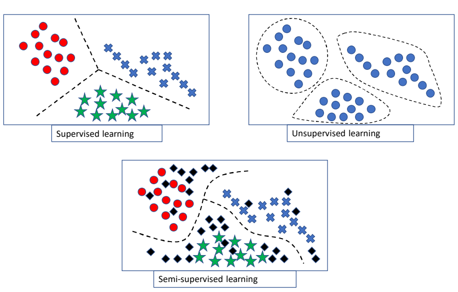
Algorithms: Types of Learning

- **Supervised Learning : Labelled Data** $(\{x_n \in R^d, y_n \in R\}_{n=1}^N)$
  - Prediction
  - Classification (discrete labels), Regression (real values)
- **Unsupervised Learning: Un-labelled Data** $(\{x_n \in R^d\}_{n=1}^N)$
  - Clustering
  - Probability distribution estimation
  - Finding association (in features)
  - Dimension reduction
- **Semi-Supervised Learning : Only a part of data is labelled**
- **Reinforcement Learning: Rewards from sequence of actions**
  - Decision making (robot, chess machine)

14

10

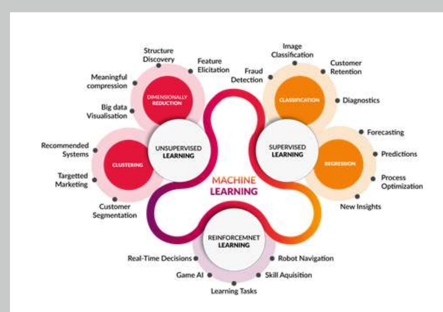
Algorithms: Types of Learning



11

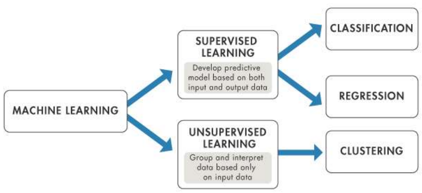
11

Types of Learning



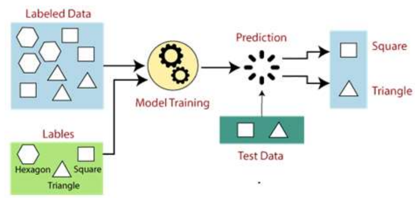
12

In this Course



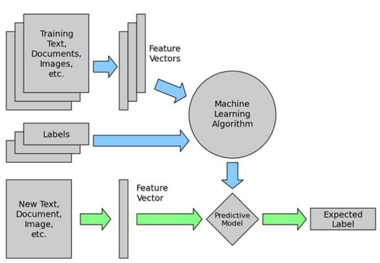
13

Supervised Learning



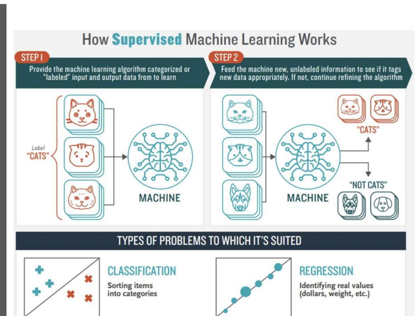
14

Supervised Learning Model



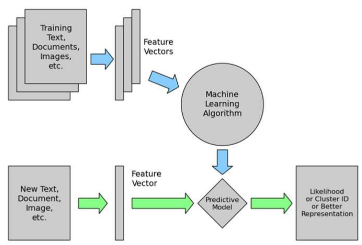
15

Example Supervised Learning



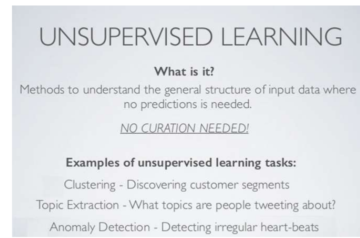
16

Unsupervised Learning Model



17

Example Unsupervised Learning


UNSUPERVISED LEARNING

What is it?

Methods to understand the general structure of input data where no predictions is needed.

NO CURATION NEEDED!

Examples of unsupervised learning tasks:

Clustering - Discovering customer segments

Topic Extraction - What topics are people tweeting about?

Anomaly Detection - Detecting irregular heart-beats

18

Machine Learning

Basic Idea: In many domains, it is difficult to hand-build a predictive model, but easy to collect lots of data; machine learning provides a way to automatically infer the predictive model from data.

Supervised Learning:

Training Data	Machine learning algorithm	Predictions
$(x^{(1)}, y^{(1)})$ $(x^{(2)}, y^{(2)})$ $(x^{(3)}, y^{(3)})$ $\vdots$	$\longrightarrow$ Hypothesis function $y^{(i)} \approx h(x^{(i)})$ $\longrightarrow$	New example x $\hat{y} = h(x)$

19

Terminology

Input features: $x^{(i)} \in \mathbb{R}^n, i = 1, \dots, m$

$$\mathrm{E.g.:}\ x^{(i)} = \begin{bmatrix} \mathrm{High\_Temperature}^{(i)} \\ \mathrm{Is\_Weekday}^{(i)} \\ 1 \end{bmatrix}$$

Outputs: $y^{(i)} \in \mathcal{Y}, i = 1, \dots, m$

$$\mathrm{E.g.:}\ y^{(i)} \in \mathbb{R} = \mathrm{Peak\_Demand}^{(i)}$$

Model parameters: $\theta \in \mathbb{R}^n$

Hypothesis function: $h_\theta : \mathbb{R}^n \to \mathcal{Y}$, predicts output given input

$$\mathrm{E.g.:}\ h_\theta(x) = \sum_{j=1}^{n} \theta_j \cdot x_j$$

20

Terminology

Loss function: $\ell: \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$, measures the difference between a prediction and an actual output

$$\text{E.g.: } \ell(\hat{y}, y) = (\hat{y} - y)^2$$

The machine learning optimization problem:

$$\underset{\theta}{\text{minimize}} \sum_{i=1}^{m} \ell(h_\theta(x^{(i)}), y^{(i)})$$

Virtually every machine learning algorithm has this form, just specify

- What is the hypothesis function?
- What is the loss function?
- How do we solve the optimization problem?

21

Example of ML Algorithms

- **Least Squares:** {linear hypothesis, squared loss, (usually) analytical solution}
- **Linear Regression:** {linear hypothesis}
- **Support Vector Machine:** {linear or kernel hypothesis, hinge loss}
- **Neural Network:** {Composed non-linear function, (usually) gradient descent}
- **Decision Tree:** {Hierarchical axis-aligned halfplanes, greedy optimization}
- **Naïve Bayes:** {Linear hypothesis, joint probability under certain independence assumptions, analytical solution}

22

Loss vs. Error vs. Cost Function

- The loss function computes the error for a single training example, while the cost function is the average of the loss functions of the entire training set.
- If we have ***m*** training data like this **{(x_1,y_1),(x_2,y_2), . .. (x_m,y_m)}**.
- $\hat{y}_i$ = output of the model for training example $\mathbf{x_i}$
- y_i = expected output/true value for training example $\mathbf{x_i}$

❑ The **loss function** $L(\hat{y}_i, y_i)$ to defines the error/difference between $\hat{y}_i$ and y_i for the single training example $\mathbf{x_i}$.

- This means, loss refers to **error in model output for an individual sample**.

❑ If we want to find loss over **all the training examples present in a training-set**, we refer to it as the **cost function** (i.e. total or average loss over all training examples).

- This is the estimate of **total error** computer for the whole training-set.

23

Loss/Error/Cost – Objective Function

The terms cost and loss functions are synonymous some people also call it error function.

The more general scenario is to define an objective function first, which we want to optimize. This objective function could be to

1. maximize the posterior probabilities (e.g., naive Bayes)
2. maximize a fitness function (genetic programming)
3. maximize the total reward/value function (reinforcement learning)
4. maximize information gain/minimize child node impurities (CART decision tree classification)
5. minimize a mean squared error cost (or loss) function (CART, decision tree regression, linear regression)
6. maximize log-likelihood or minimize cross-entropy loss (or cost) function (ANN), minimize hinge loss (support vector machine)

24

Some General Applications of ML

Classification/ Regression/ Clustering/Association

- **Image Processing**
- **Robotics**
- **Data Mining**
- **Recommender Systems**
- **Game Playing**
- **Text processing**
- **Health care**

Source: cbinsights, Quora, dbta, Futurism, Wikipedia

25

25

References

- Trevor Hastie, Robert Tibshirani, Jerome Friedman; "*The Elements of Statistical Learning: Data Mining, Inference, and Prediction*"; Springer.
- Tom Mitchell; "*Machine Learning*", McGraw Hill
- Russell, Stuart J.; Norvig, Peter; "*Artificial Intelligence: A Modern Approach*"; Prentice Hall.

26

26

27

To be continued in the next session…..

28