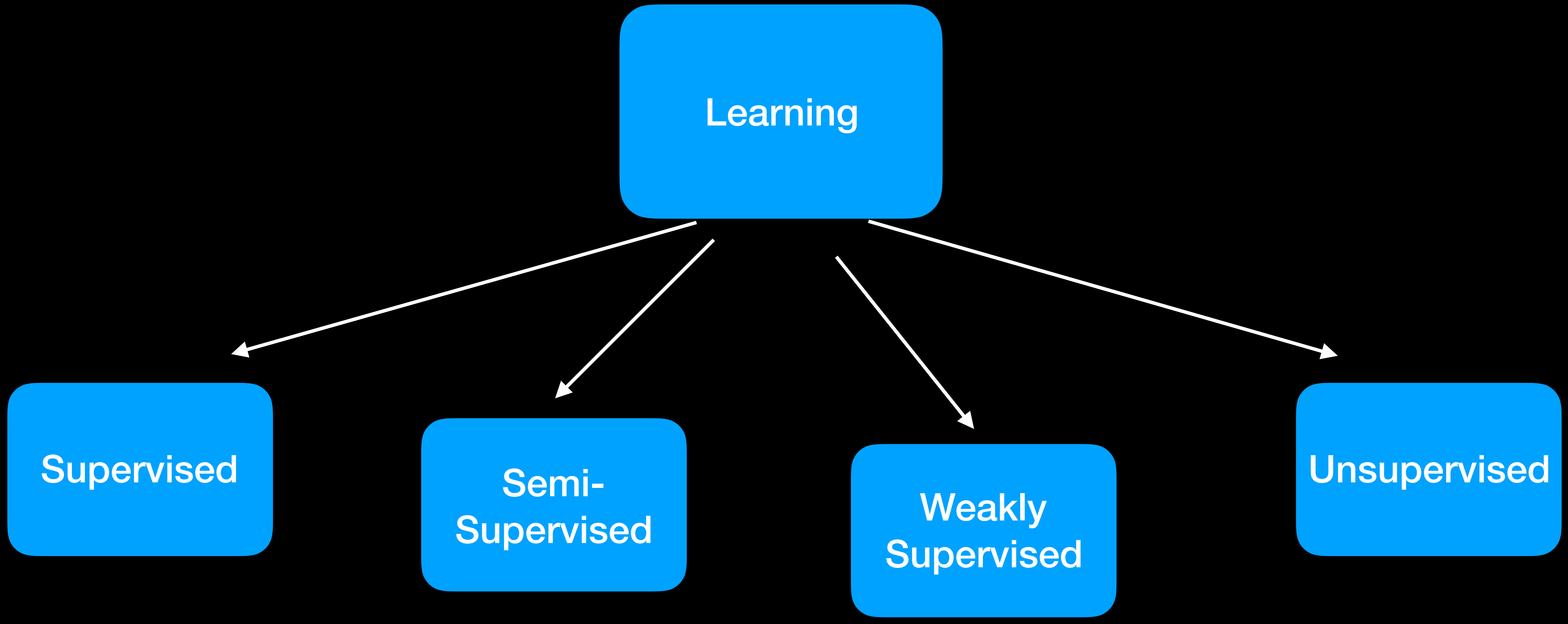
18th July 2019

Lecture 2

Machine Learning : Foundations and Applications

Transcribed by Vibhu N. Deo (vibhund@gmail.com)


Learning
Supervised
Semi-Supervised
Weakly Supervised
Unsupervised

Supervised Learning

- Comprises of two phases:
  - Training Phase
    - Input-Output Pairs are provided
    - A function $f(\overrightarrow{X_i})$ is generated that links the input to the output
  - Testing Phase
    - An input is provided to the function $f(\overrightarrow{X_i})$ which finds the corresponding output $\overrightarrow{y_i}$

Supervised Learning

- N = number of training examples
- The "input" is a D dimensional vector $\overrightarrow{X_i}$ where $i = 1,2,\ldots,N$
- $\overrightarrow{X_i}$ is a vector of Real number components. It is also called the **feature vector**.
- The "output" is also called the **label vector.** It is denoted by $\overrightarrow{y_i}$ where $i = 1,2,\ldots,N$

- The label vector can belong to any of the following spaces:
  - Binary Space
    - E.g. [‘cat’, ‘dog’], [0,1]
    - Occurs in Classification problems
  - Discrete Space
    - E.g. [‘cat’, ‘dog’, ‘bird’,…,(k values)]
    - Occurs in Classification problems (k - space classification)
  - Real Numbers $\mathbb{R}$
    - Occurs in Regression problems
  - Structured Output
    - Structured depend on the problem or the context
    - Occurs in Structured Prediction problems

Classification

- Input = $[(\overrightarrow{X_i}, \overrightarrow{y_i})]_N$,where $\overrightarrow{X_i}$ is a D dimensional feature vector and $\overrightarrow{y_i}$ is the label vector
- Training:
  - We try to find a function f such that $f : \mathcal{X} \rightarrow \mathcal{Y}$, where $\mathcal{X}$ is the feature space and $\mathcal{Y}$ is the label space
  - Depending on the algorithm used, f is created.
- Testing:
  - We apply f on $\overrightarrow{X}_{test} \in \mathcal{X}$
  - We get $f(\overrightarrow{X}_{test}) = \overrightarrow{y}_{pred}$

Loss Function

- Used to estimate how much “loss” we have incurred from our prediction in comparison to the actual value

- The lower the value of the Loss Function L, the greater the correctness of the prediction.

- $L = L(\overrightarrow{y}_{pred}, \overrightarrow{y}_{test})$

Types of Loss Functions

- Binary Loss Function : (0,1)

- Confusion Matrix or Loss Matrix

Prediction/ Test	0	1
0	0	Small value
1	Large value	0

- This is an example for a 2×2 matrix. There can be a $k \times k$ matrix also for a discrete space of k elements.

- In case of Regressions, the label space consists of Real numbers.
- Hence, a Binary Loss Function or Confusion Matrices are not applicable.
- Instead, we use the squared errors:
  $L(y_{pred}, y_{test}) = (y_{pred} - y_{test})^2$
- We use squared error because these values are always positive.
- Test Loss = $\sum_{i=1}^{N} L(y_i^{pred}, y_i^{test})$

- Questions
  - How to convert an input into a feature vector?
    - There is no standard answer. It is dependant on the context of the problem.
  - How to find f?
    - There are some algorithms present to help us find a suitable f.

Thank you