Semi-supervise Learning

Ramesh Ragala
VIT Chennai

Semi-supervised Learning

- Semi-supervised learning (SSL) is a class of machine learning techniques that make use of both labeled and unlabeled data for training.
- Unsupervised Learning
  - Let $X = (x_1, x_2, \ldots x_n)$ be a set of n examples or points, where $x_i \in \mathcal{X}$ for all $i \in [n] = \{1, 2, \ldots, n\}$.
  - it is assumed that the points are drawn i.i.d. (independently and identically distributed) from a common distribution on $\mathcal{X}$.
  - The goal of unsupervised learning is to find interesting structure in the data X.
  - It has been argued that the problem of unsupervised learning is fundamentally that of estimating a density which is likely to have generated X.

Semi-supervised Learning

- supervised Learning
  - The goal is to learn a mapping from x to y, given a training set made of pairs (x_i,y_i).
  - Here, the $y_i \in Y$ are called the labels or targets of the examples x_i → If the labels are numbers Y denotes the column vector of labels.
  - The pairs (x_i, y_i) are sampled i.i.d. from some distribution which here ranges over X × Y. → This task is well defined, since a mapping can be evaluated through its predictive performance on test examples
  - it is assumed that the points are drawn i.i.d. (independently and identically distributed) from a common distribution on $\mathcal{X}$.
  - When $y \in R$ or R^d (i.e. when the labels are continuous), the task is called regression.
  - There are two families of algorithms for supervised learning.
    - Generative algorithms → try to model the class-conditional density by some unsupervised learning procedure. A predictive density can
    - then be inferred by applying Bayes theorem

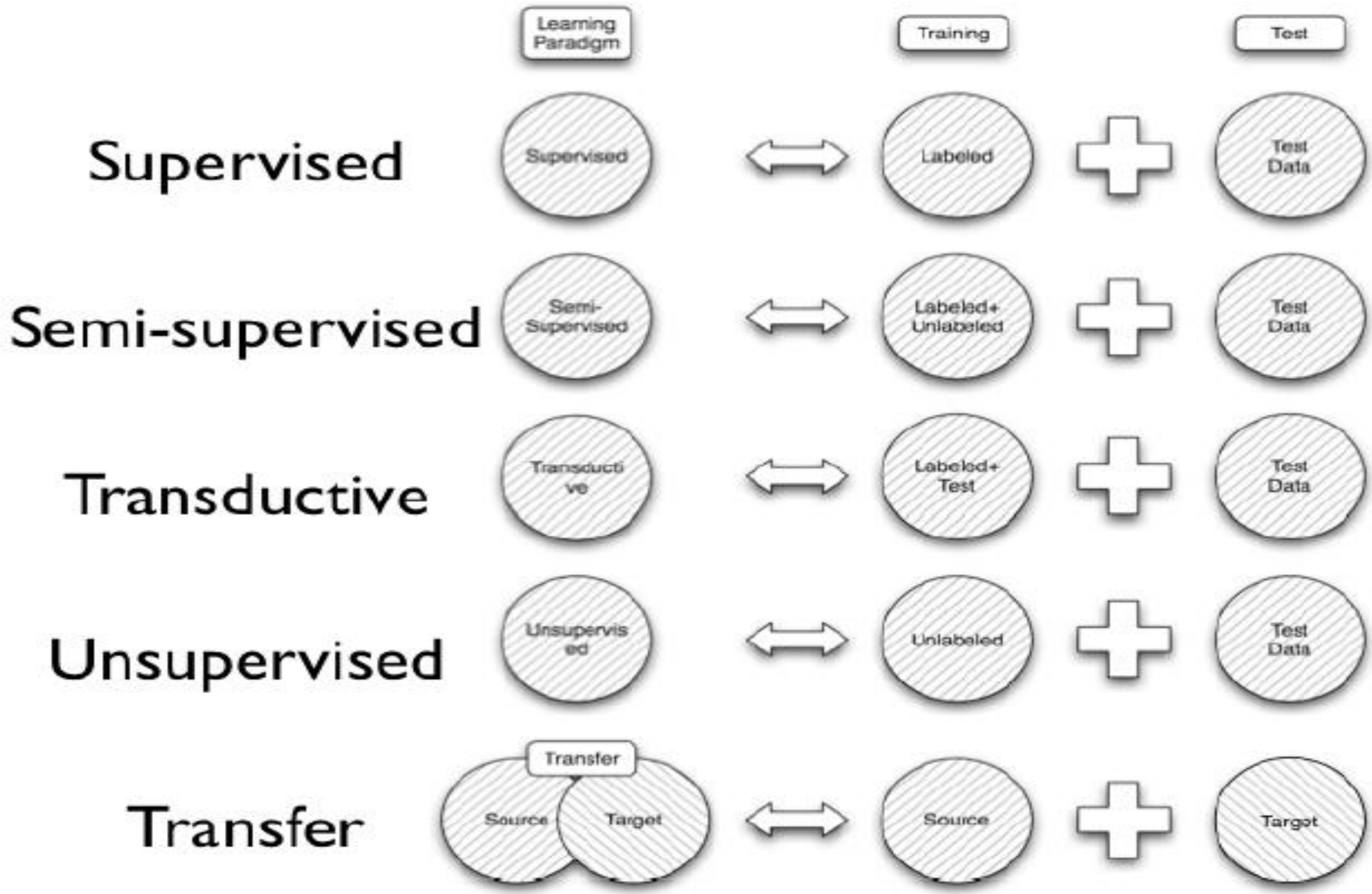
Semi-supervised Learning

- supervised Learning
- There are two families of algorithms for supervised learning.
  - Generative algorithms → try to model the class-conditional density p(x/y) by some unsupervised learning procedure. A predictive density can then be inferred by applying Bayes theorem:

$$p(y|x) = \frac{p(x|y)p(y)}{\int_y p(x|y)p(y)dy}.$$

  - Discriminative algorithms → do not try to estimate how the x_i have been generated, but instead concentrate on estimating p(y/x). Some discriminative methods even limit themselves to modeling whether p(y/x) is greater than or less than 0.5. Example SVM

Semi-supervised Learning



Semi-supervised Learning

- Semi-supervised learning falls between
  - Unsupervised learning (without any labeled training data) and
  - Supervised learning (with completely training data)
- Learn predictive tasks
  - Uses both labeled data and unlabeled data
  - Small amount of labeled data
  - Large amount of unlabeled data
- The dataset $X = (x_i)_{i \in [n]}$ can be divided into two parts:
  - The points $X_l = (x_1, x_2, \ldots, x_l)$ for which labels $Y_l = (y_1, y_2, \ldots y_l)$ are provided.
  - The points $X_u = (x_{l+1}, x_{l+2}, \ldots x_{l+u})$ does not know the labels
  - This is a standard semi-supervised learning
- Semi-supervised learning with constraints
  - Partial supervision is possible
  - Example: these points have the same target.

Semi-supervised Learning

- Two types of learning will be used in Semi-supervised learning and sometime semi-supervised learning may refer either of
  - Transductive learning
  - Inductive learning
- Transductive learning
  - The idea of transduction is to perform predictions only for the test points. i.e it is used to infer the correct labels for the given unlabeled data (x_{l+1}, x_{l+2}, … x_{l+u}) only.

Given $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{l}$ and $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$, learn a function $f : \mathcal{X}^{l+u} \longrightarrow \mathcal{Y}^{l+u}$ so that f is expected to be a good predictor on the unlabeled data $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$.

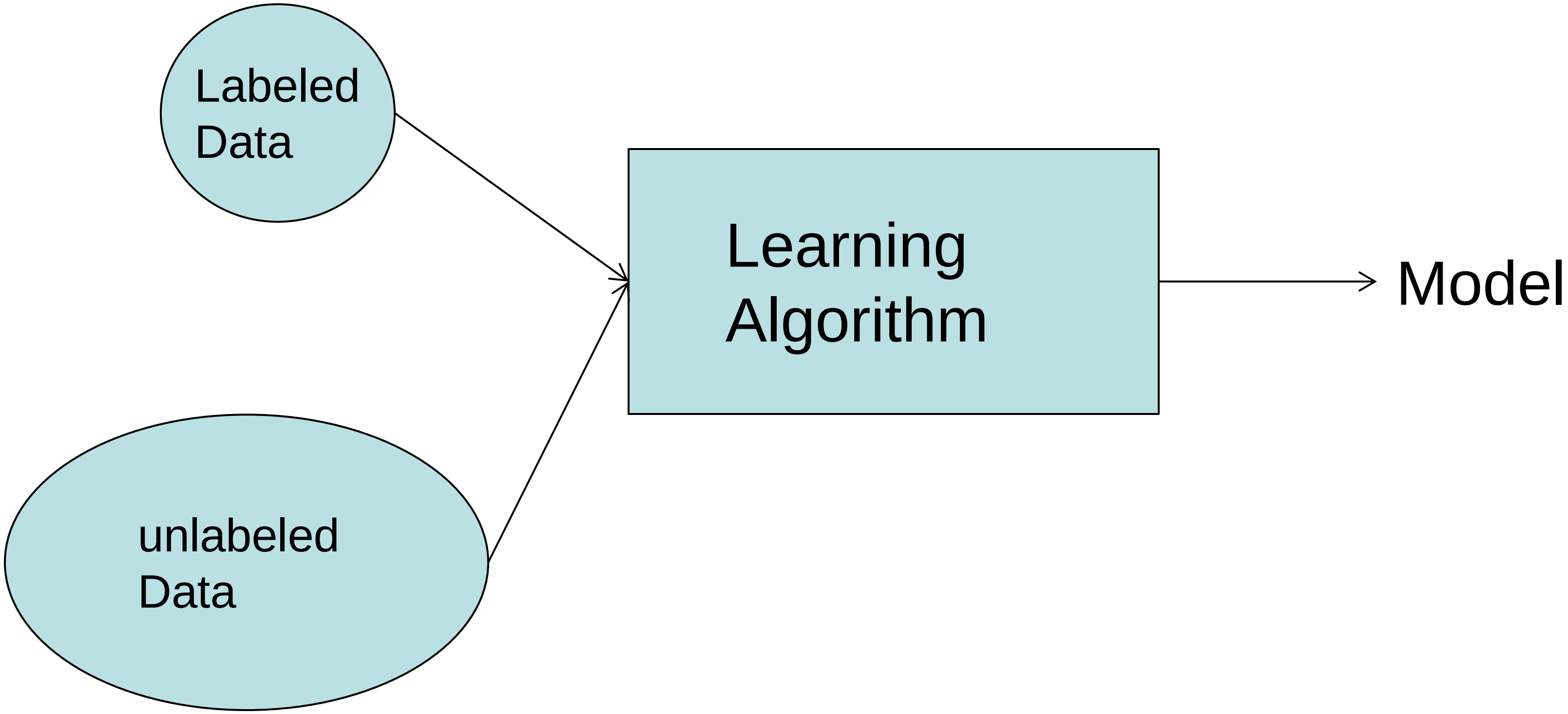
- Inductive learning
  - the goal is to output a prediction function which is defined on the entire space X. i.e. it is used to infer the correct mapping from X to Y.

Given $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{l}$ and $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$, learn a function $f : \mathcal{X} \longrightarrow \mathcal{Y}$ so that f is expected to be a good predictor on future data.

Semi-supervised Learning

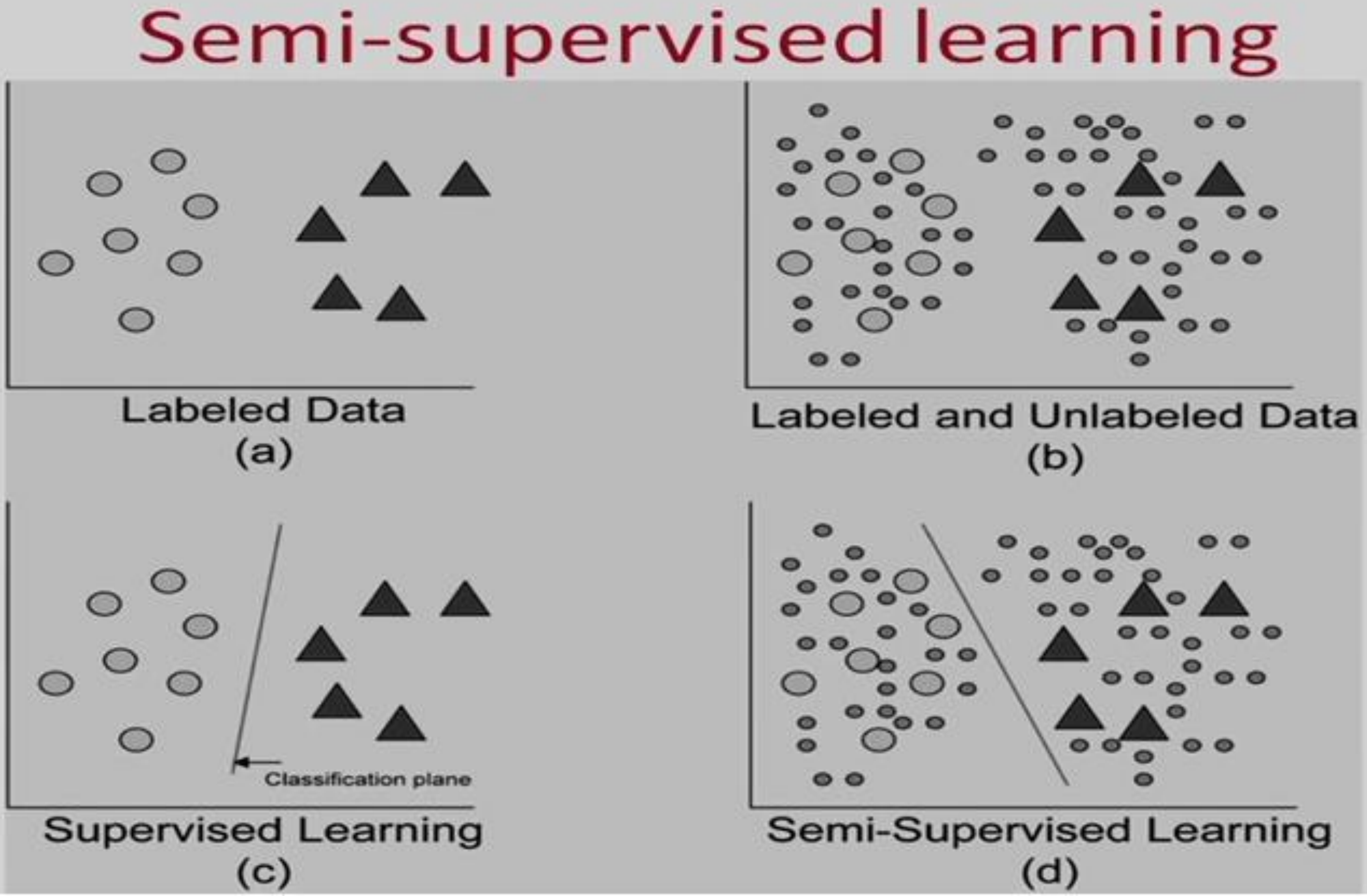
- Examples of Semi-supervised learning
  - Self-training
  - Co-training
- Applications of Semi-supervised Learning
  - Speech analysis
  - Telephone conversation transcription
    - 400 hours annotation time for each hour of speech
  - Protein sequence classification
  - Web page classification

Semi-supervised Learning



Simple architecture of Semi-supervised learning

Semi-supervised Learning

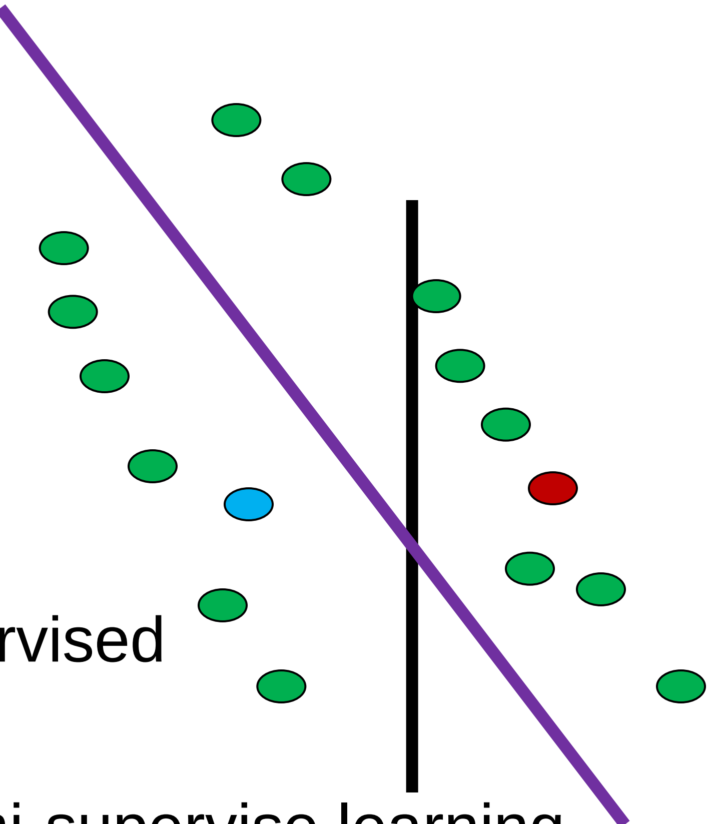


Need of Semi-supervised Learning

- Labeled data is costly for many applications
- The acquisition of labeled data for learning problem often requires a skilled human agent or a physical experiments
- Examples:
  - Speech Analysis
  - Classification of web based text
- Unlabeled data is not expensive and able to get large quantity also
- By using these combination, it can produce considerable improvement in learning accuracy

Semi-supervised Learning

Supervised Learning



Class – 1 sample

Class – 2 sample

Decision boundary using supervised

Unlabeled sample

Decision boundary using semi-supervise learning

Semi-supervised Learning

- The classes of semi-supervised learning methods
  - Generative Models
  - Low – Density Separation
  - Graph – Based Methods
  - Change of Representation
  - Self-training
  - Co-training