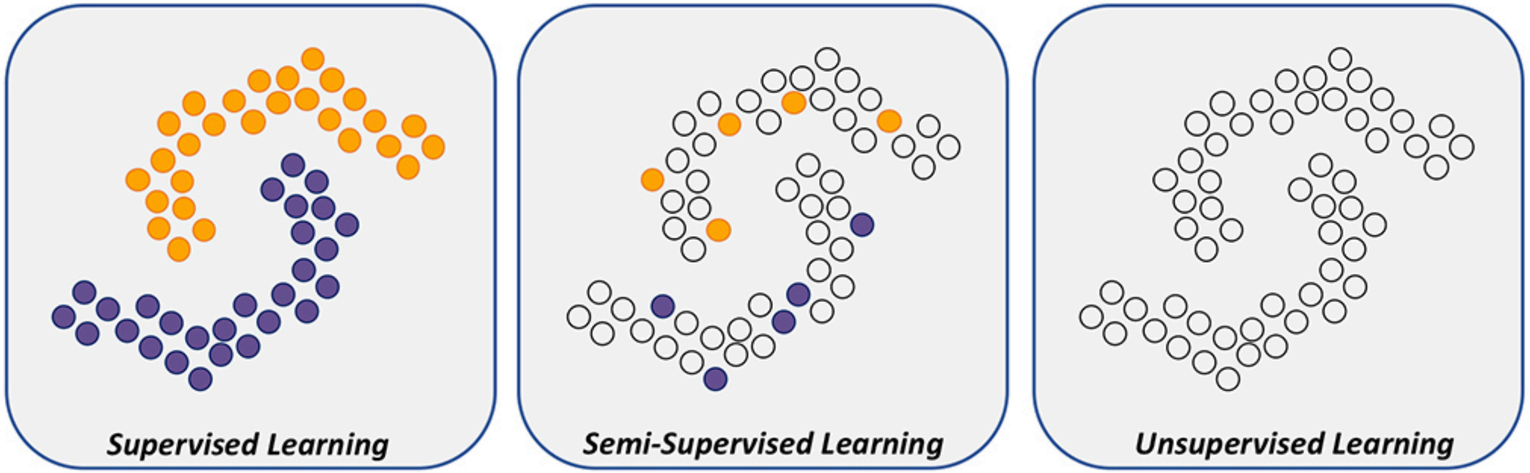
Semi-Supervised Learning

Basic concept of Semi-Supervised Learning

Hyungjin Kim

What does ‘Semi’ mean?

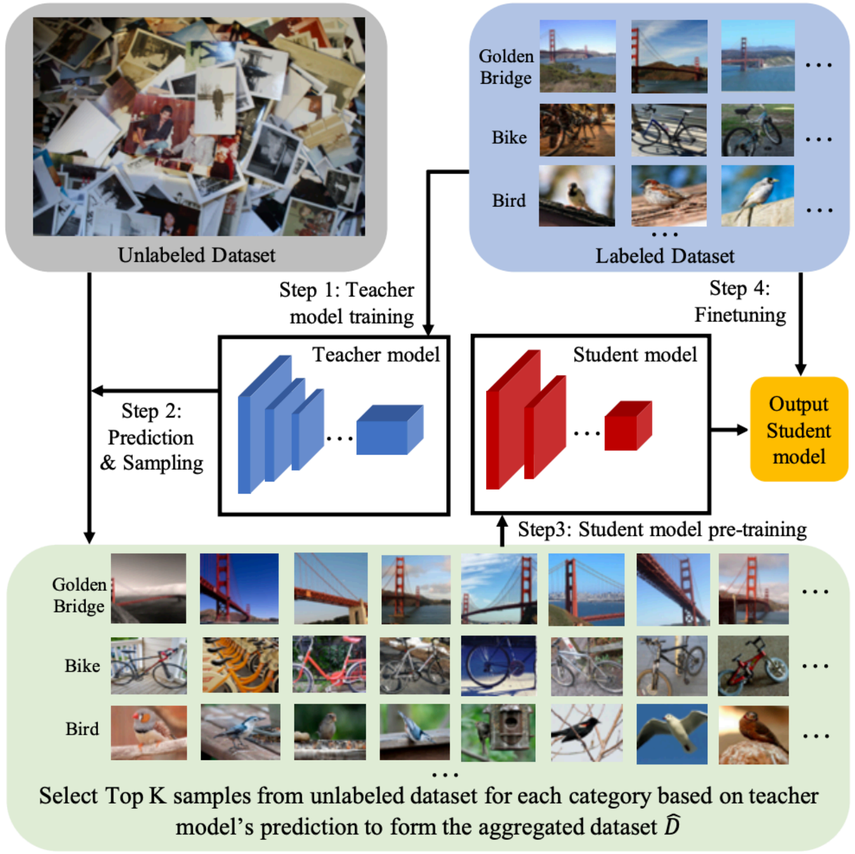
Supervised Learning + Unsupervised Learning



An approach to machine learning that combines
a small amount of labeled data with a large amount of unlabeled data during training

Why do we need to use SSL?

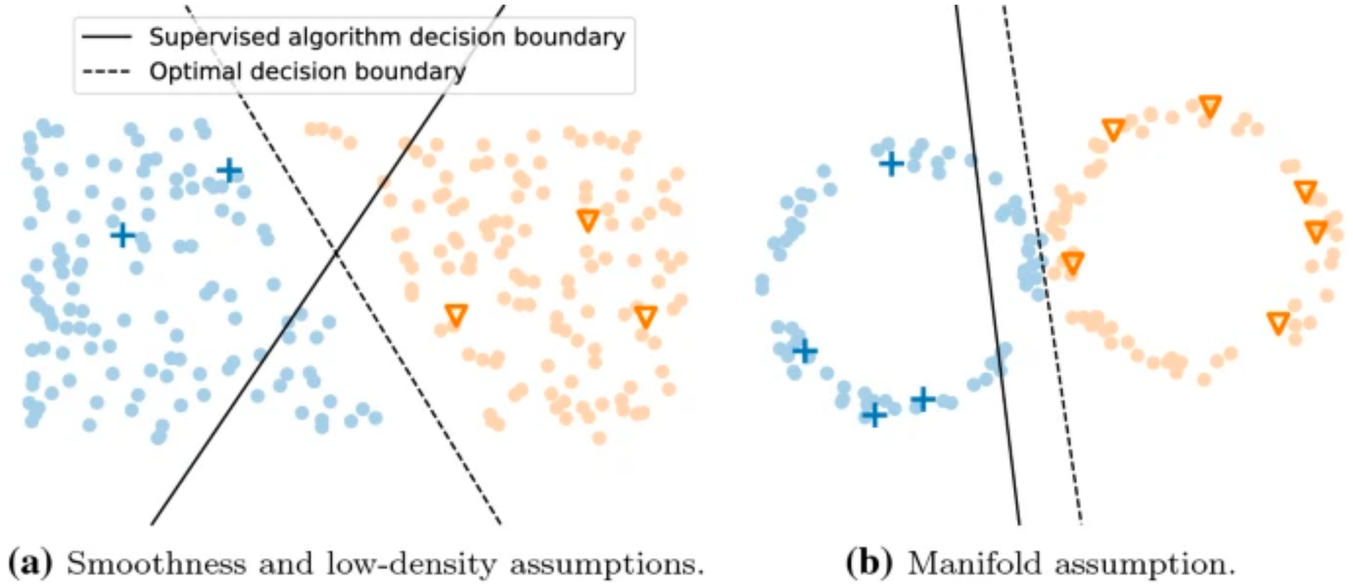
Advantages of Semi-Supervised Learning



- Labelled data is expensive and difficult to get while unlabelled is abundant and cheap.
- It improves the model robustness by having more precise decision boundary.
- It can be used for both ML & DL algorithms (NLP, Vision etc).

Assumptions of SSL

Smoothness



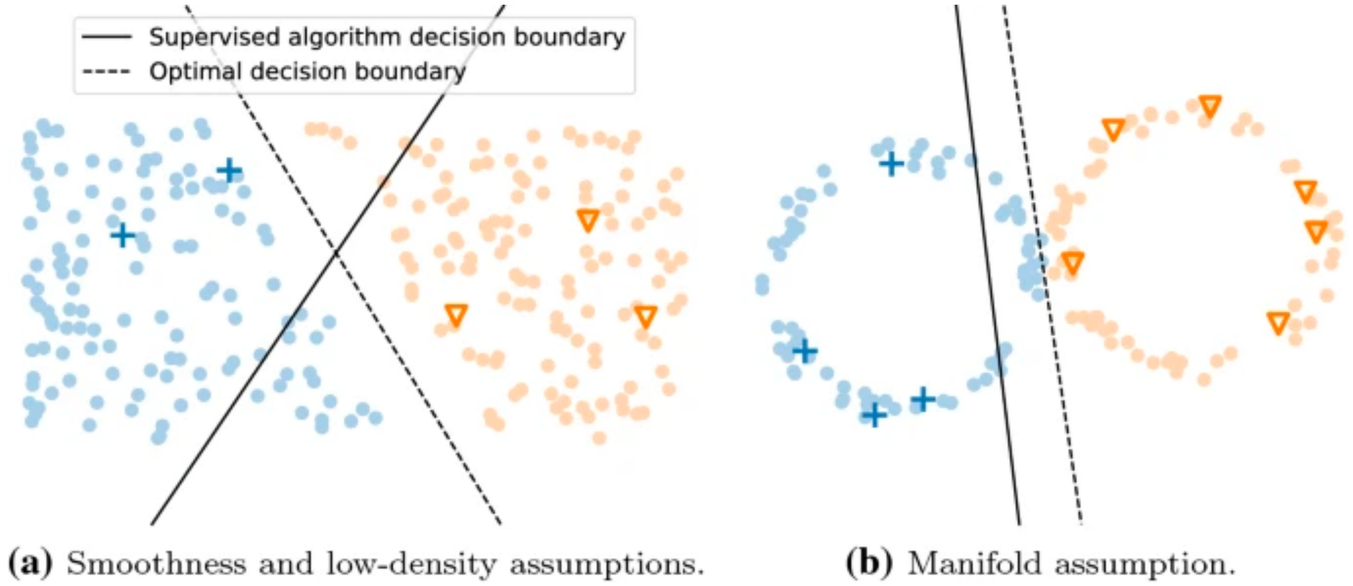
(a) Smoothness and low-density assumptions. **(b)** Manifold assumption.

Smoothness Assumption:

If two samples x1 and x2 are close in the input space, their labels y1 and y2 should be the same

Assumptions of SSL

Low-density



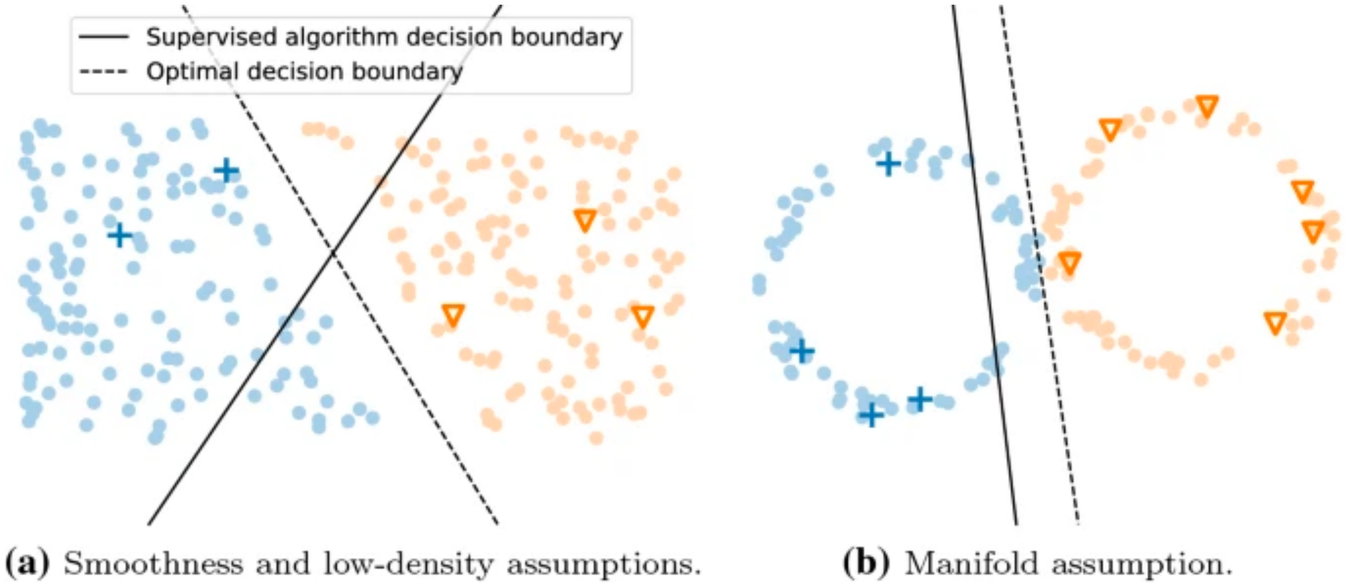
(a) Smoothness and low-density assumptions. **(b)** Manifold assumption.

Low-density Assumption:

The decision boundary should not pass through high-density areas in the input space

Assumptions of SSL

Manifold

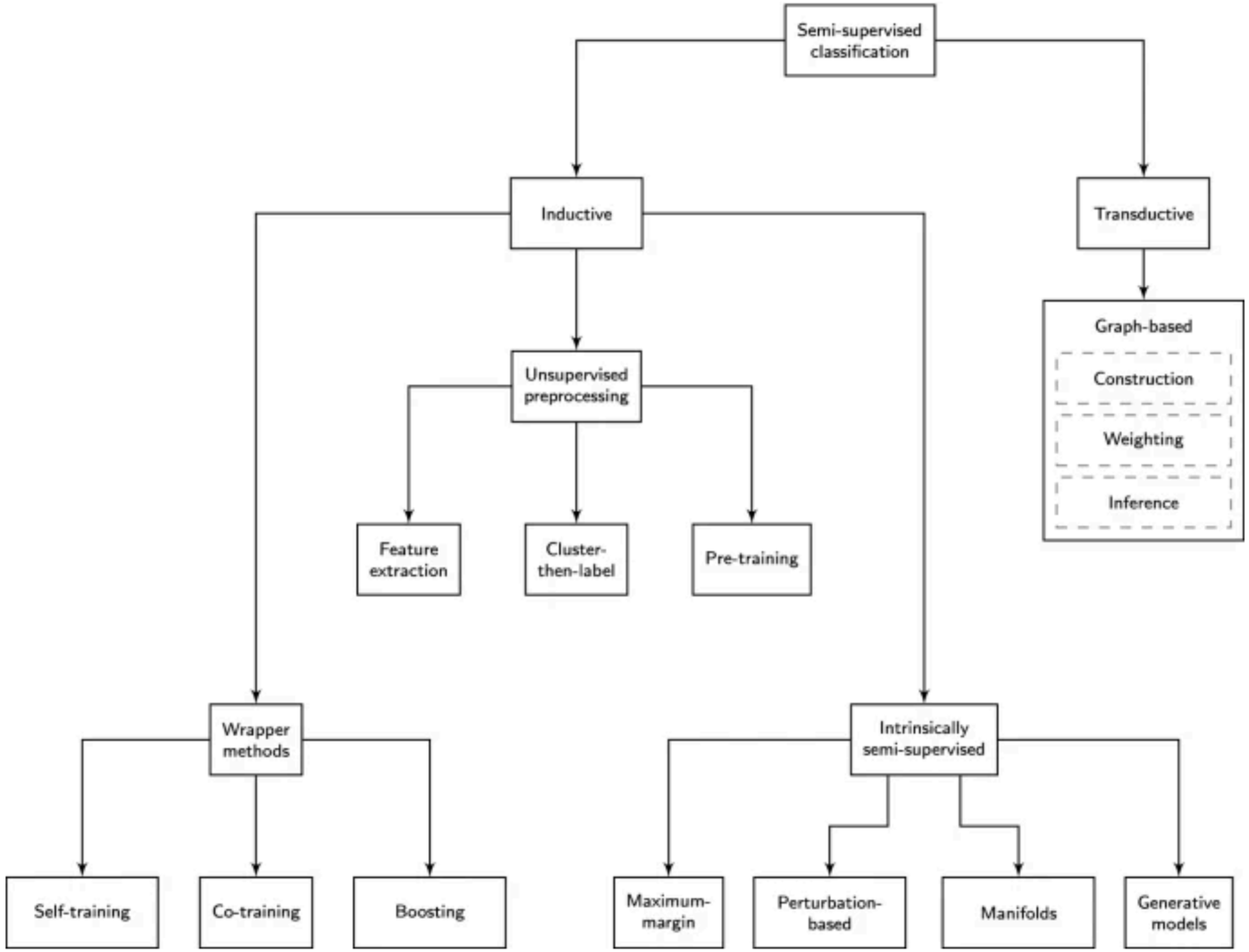


(a) Smoothness and low-density assumptions. **(b)** Manifold assumption.

Manifold Assumption:

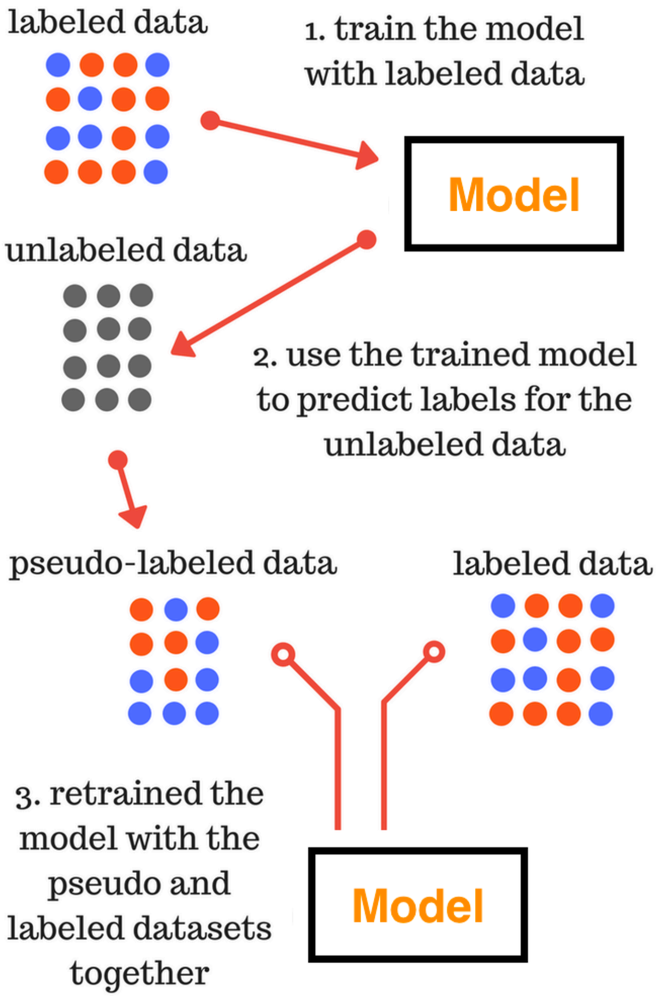
Data points on the same low-dimensional manifold should have same label

Taxonomy of Semi-supervised classification



Example of SSL labeling techniques

Pseudo-Labeling



Widely used because it is very simple

If the amount of supervised data is not enough, the performance may not be good

References

- https://blog.est.ai/2020/11/ssl/
- https://machinelearningmastery.com/semi-supervised-learning-with-label-propagation/
- https://ai.googleblog.com/2019/07/advancing-semi-supervised-learning-with.html
- https://en.wikipedia.org/wiki/Semi-supervised_learning
- https://jiwunghyun.medium.com/semi-supervised-learning-%EC%A0%95%EB%A6%AC-a7ed58a8f023

E.O.D.