University of Roehampton London

Machine Learning

Dr Changjiang He, Dr Kuo-Ming Chao
Computer Science| School of Art
University of Roehampton

What is Semi-supervised Learning?

University of Roehampton London

- **Semi-supervised learning** is a learning problem that involves a small number of labelled examples and a large number of unlabelled examples.
- Learning problems of this type are challenging as neither supervised nor unsupervised learning algorithms are able to make effective use of the mixtures of labelled and untellable data.
- As such, specialized semis-supervised learning algorithms are required.

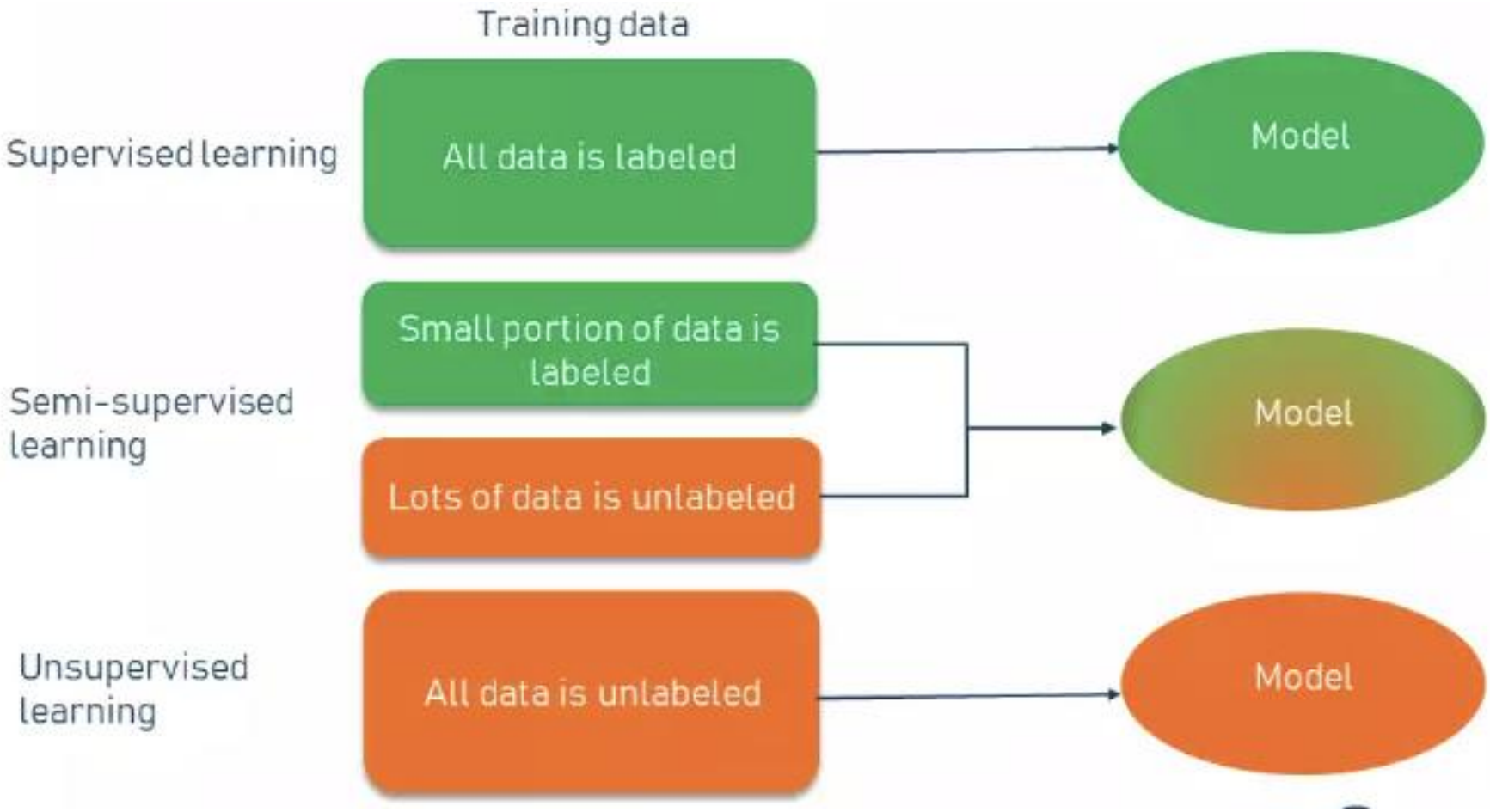
University of Roehampton London

Lesson 8.1

Semi-supervised Learning

What is Semi-supervised Learning?

University of Roehampton London



How Semi-supervised Learning Works?

University of Roehampton London

- Instead of adding tags to the entire dataset, we go through and hand-label just a small part of the data and use it to train a model, which then is applied to the ocean of unlabelled data.

- One of the simplest examples of semi-supervised learning, in general, is self-training.

- **Self-training** is the procedure in which you can take any supervised method for classification or regression and modify it to work in a semi-supervised manner, taking advantage of labelled and unlabelled data.

Applications

University of Roehampton London

- With the amount of data constantly growing by leaps and bounds, there's no way for it to be labelled in a timely fashion.
- Think of an active TikTok user that uploads up to 20 videos per day on average. And there are 1 billion active users.
- In such a scenario, semi-supervised learning can boast of a wide array of use cases from image and speech recognition to web content and text document classification.

Applications

University of Roehampton London

- Speech recognition.
  Labelling audio is a very resource- and time-intensive task, so semi-supervised learning can be used to overcome the challenges and provide better performance.

- Web content classification.
  With billions of websites presenting all sorts of content out there, classification would take a huge team of human resources to organize information on web pages by adding corresponding labels. The variations of semi-supervised learning are used to annotate web content and classify it accordingly to improve user experience.

Applications

University of Roehampton London

- Text document classification
  Another example of when semi-supervised learning can be used successfully is in the building of a text document classifier. Here, the method is effective because it is really difficult for human annotators to read through multiple word-heavy texts to assign a basic label, like a type or genre.