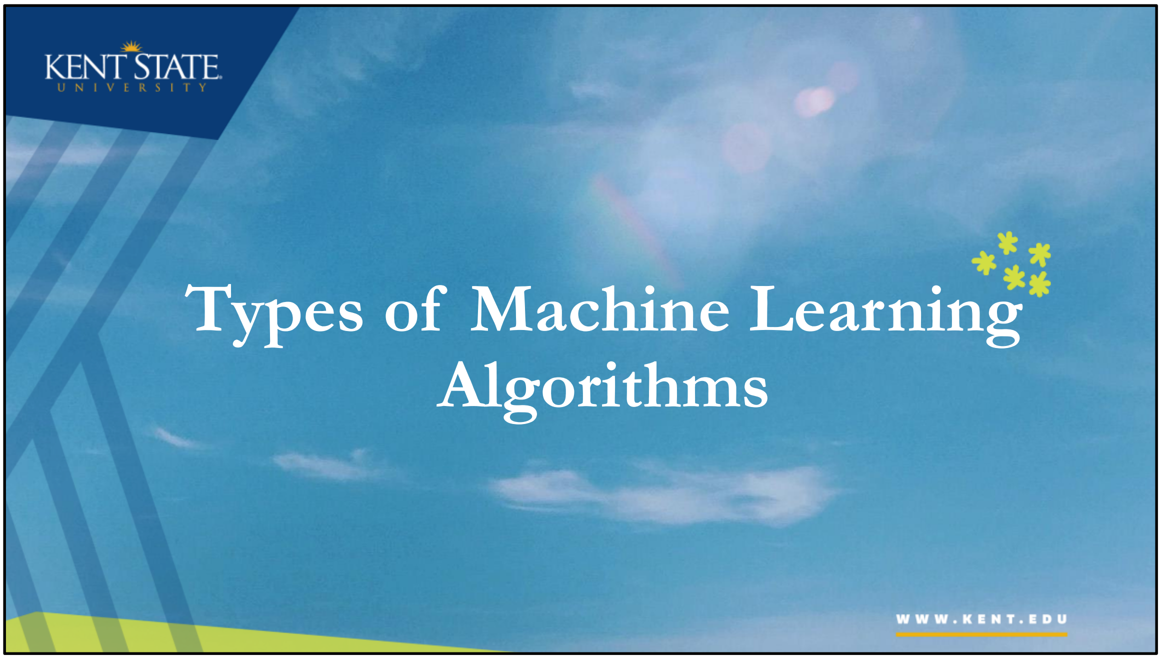


In this lecture, we discuss the various types of ML algorithms.

Types of Machine Learning Models

- Machine learning algorithms are organized into taxonomy, based on the desired outcome of the algorithm. Common algorithm types include:
  - Supervised learning
  - Unsupervised learning
  - Semi-supervised learning
  - Reinforcement learning

WWW.KENT.EDU

Machine-learning algorithms generally fall into four broad categories:

Supervised learning is by far the most common case. It consists of learning to map input data to known targets given a set of examples (often annotated by humans). Almost all applications in the spotlight these days belong in this category, such as optical character recognition, speech recognition, image classification, and language translation.

Unsupervised learning consists of finding interesting transformations of the input data without the help of any targets, for the purposes of data visualization, data compression, or data denoising, or to better understand the correlations present in the data at hand.

This is a specific instance of supervised learning, but it’s different enough that it deserves its own category. Self-supervised learning is supervised learning without human-annotated labels—you can think of it as supervised learning without any humans in the loop. There are still labels involved (because the learning has to be supervised by something), but they’re generated from the input data, typically using a heuristic algorithm. *Autoencoders* are a well-known example of self-supervised learning, where the generated targets are the input, unmodified.

Long overlooked, Reinforcement learning recently started to get a lot of attention after Google DeepMind successfully applied it to learning to play Atari games (and, later, learning to play Go at the highest level). In reinforcement learning, an *agent* receives information about its environment and learns to choose actions that will maximize some reward. For instance, a neural network that “looks” at a video- game screen and outputs game actions in order to maximize its score can be trained via reinforcement learning.

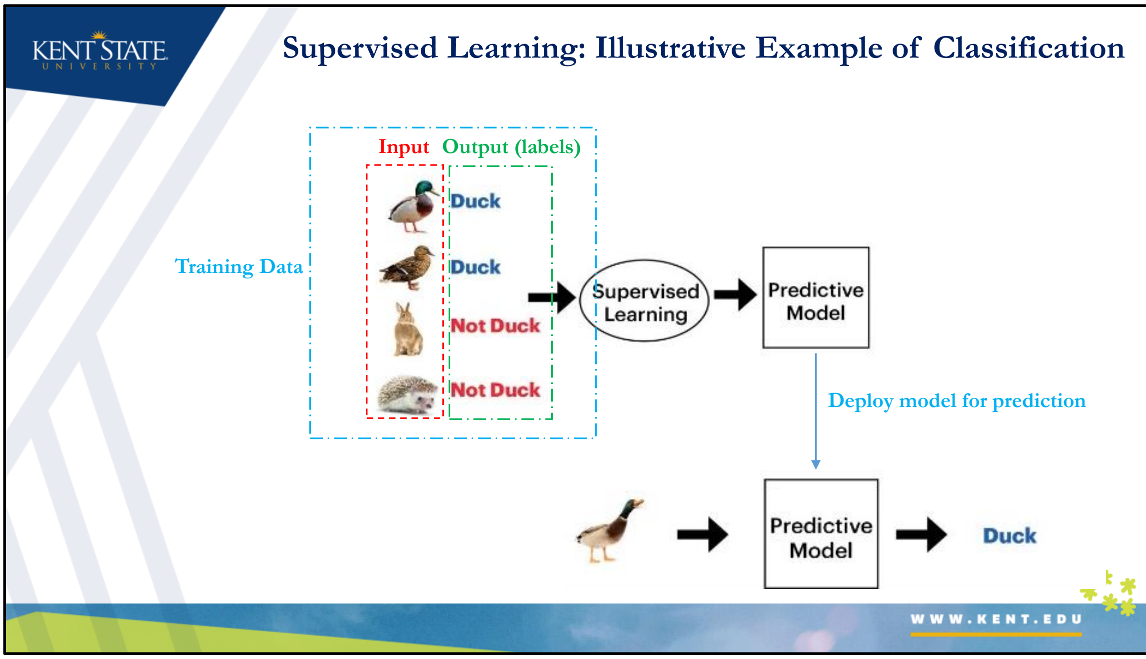
KENT STATE UNIVERSITY

Supervised learning

- In supervised learning, the algorithm generates a function that maps inputs to desired outputs. (i.e. function approximation).
- For some examples (i.e. training data), the correct outputs (i.e. target) are given to the algorithm. The algorithm will then try to find the best function that maps the input to the outputs during the training (i.e. learning) phase.
- Once the model is trained, it can be used to for making predictions for input values with unknown outputs.
- Supervised models are the most important types of machine learning models with wide business applications.

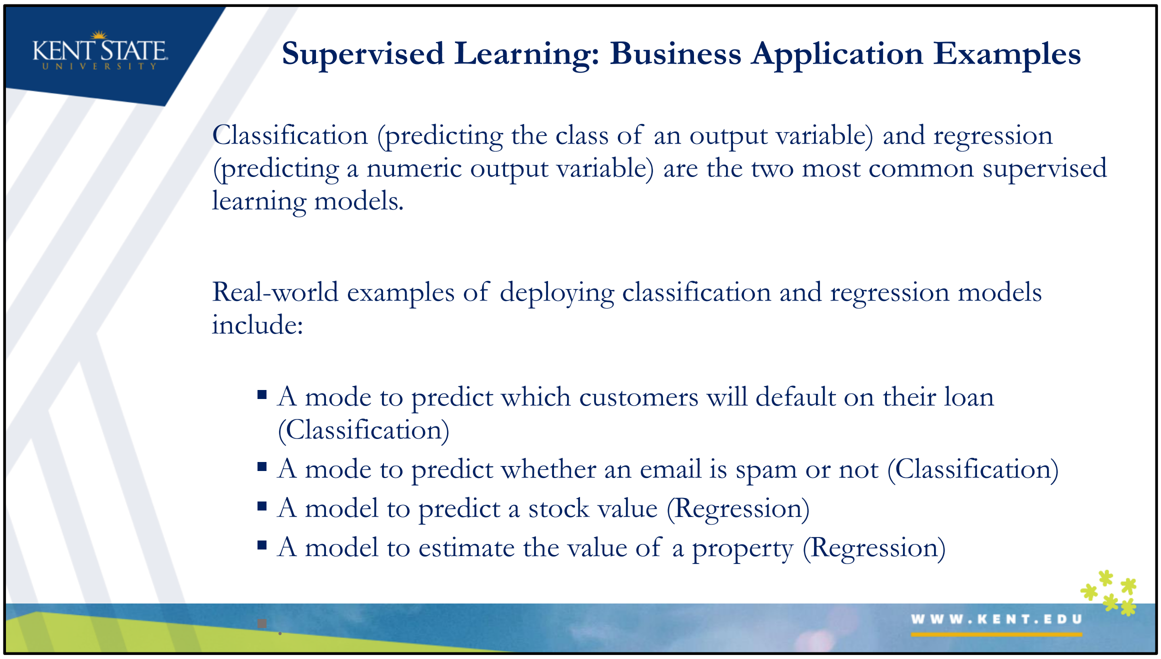
WWW.KENT.EDU

Classification, and regression are all examples of supervised learning. Training a model, i.e., of finding the best function to map the input to the output is called learning. Once a model has been trained, it can then be used for prediction. This class of models are the most common type encountered in real-world applications.



Here is a simple example to illustrate supervised learning. Note the following:

1. Input data that is labelled (either by humans or otherwise).
2. A model. We need to define how performance is measured. For a classification problem, can you think of what the objective function performance measure might be?
3. Once a model is trained, it is then used for prediction. Question: Will the best trained model also be the best predictive model?



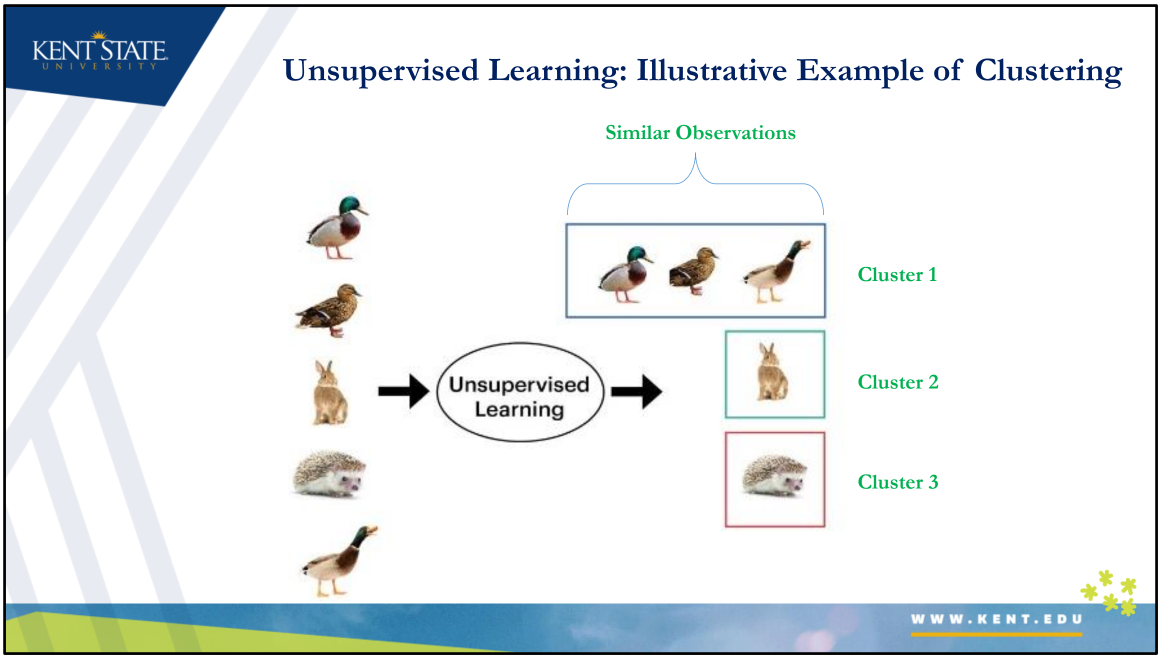
We will see many examples of classification and regression in this program. While common, they still pose challenges to developing good predictive models.

Question: List out some possible performance measures for both classification and regression type problems.

Unsupervised learning

- In supervised learning, only input observations are used in the modeling process (i.e. labeled examples are not available).
- In unsupervised learning, the algorithms are left to themselves to discover interesting structures in the data.
- The most common unsupervised learning method is cluster analysis, which is used for exploratory data analysis to find hidden patterns or grouping in data.

Unsupervised learning is the bread and butter of data analytics, and it’s often a necessary step in better understanding a dataset before attempting to solve a supervised-learning problem. *Dimensionality reduction* and *clustering* are well-known categories of unsupervised learning.

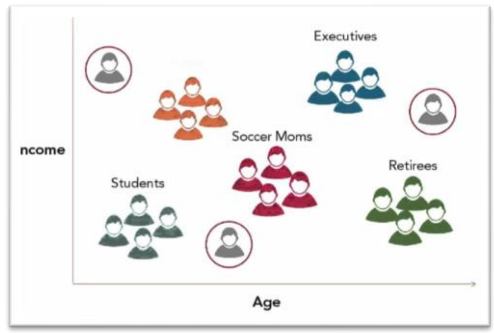


This example illustrates cluster analysis. Note that it is possible to have several types of clusters, the example only shows one, depending on what attributes are used to do the cluster analysis. For example, you might group cities like NY, Tokyo, or Mumbai based on population into one cluster, or they may belong to different clusters if grouped by geography.

KENT STATE
UNIVERSITY

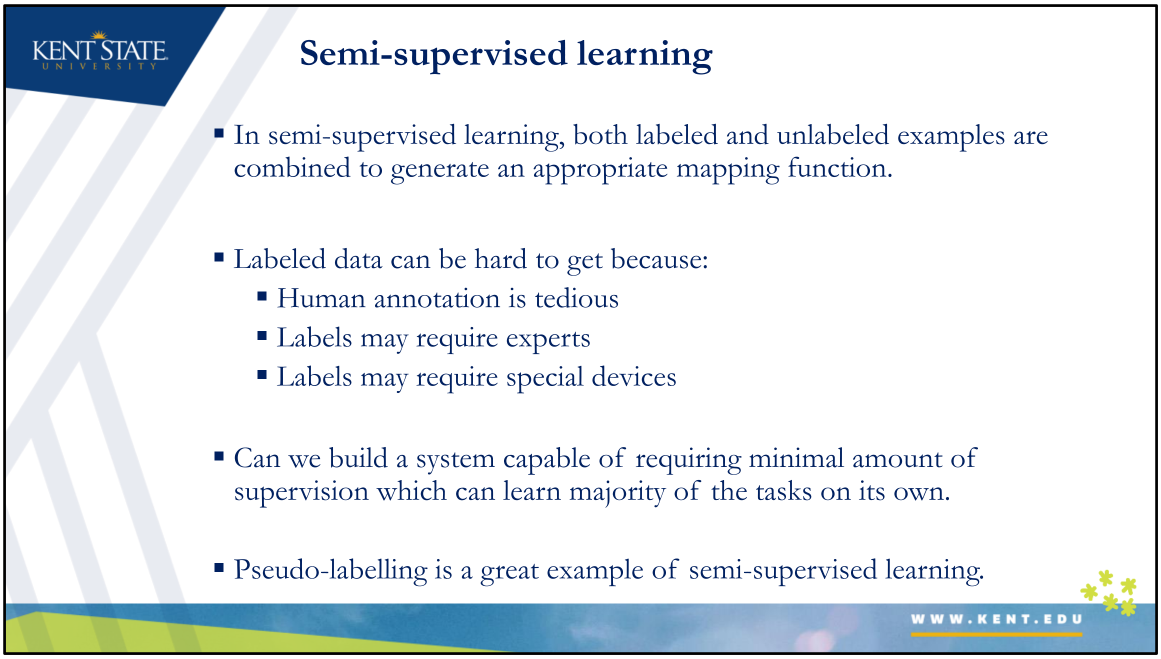
Unsupervised Learning: Business Application Examples

- Real-world examples of unsupervised machine learning models include:
  - Customer segmentation according to their demographic attributes
  - Recommendation systems
  - Anomaly detection



WWW.KENT.EDU

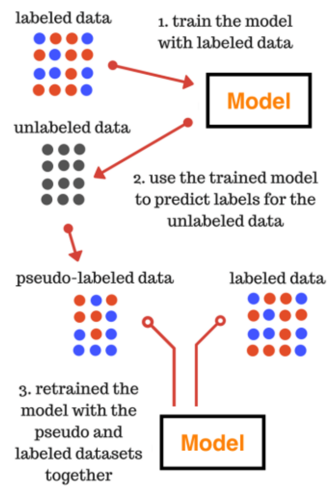
These are all well-known examples of cluster analysis. When doing such analysis, we not only provide data, but also the attributes that we want the algorithm to consider for clustering. As we will see later, there are two basic approaches to clustering: Hierarchical clustering, and k-means clustering.



The main difference between supervised and semi-supervised learning is that here the labels are not human generated. For instance, *autoencoders* are a well-known example of self-supervised learning, where the generated targets are the input, unmodified. In the same way, trying to predict the next frame in a video, given past frames, or the next word in a text, given previous words, are instances of self-supervised learning. Note that the distinction between supervised, self-supervised, and unsupervised learning can be blurry sometimes—these categories are more of a continuum without solid borders. Self- supervised learning can be reinterpreted as either supervised or unsupervised learning, depending on whether you pay attention to the learning mechanism or to the context of its application.

Pseudo-labelling

- Use the labeled data to train a model
- Use the model to predict the label of unlabeled data
- Build a model on the combined dataset

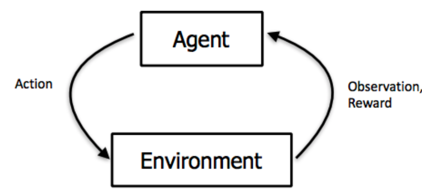


In this example, we use a model on labelled data to label unlabeled data. We then build a model on the combined data. This is similar to how Google Photos might create an album based on a person's face. If you haven't tried it yet, you should.

KENT STATE UNIVERSITY

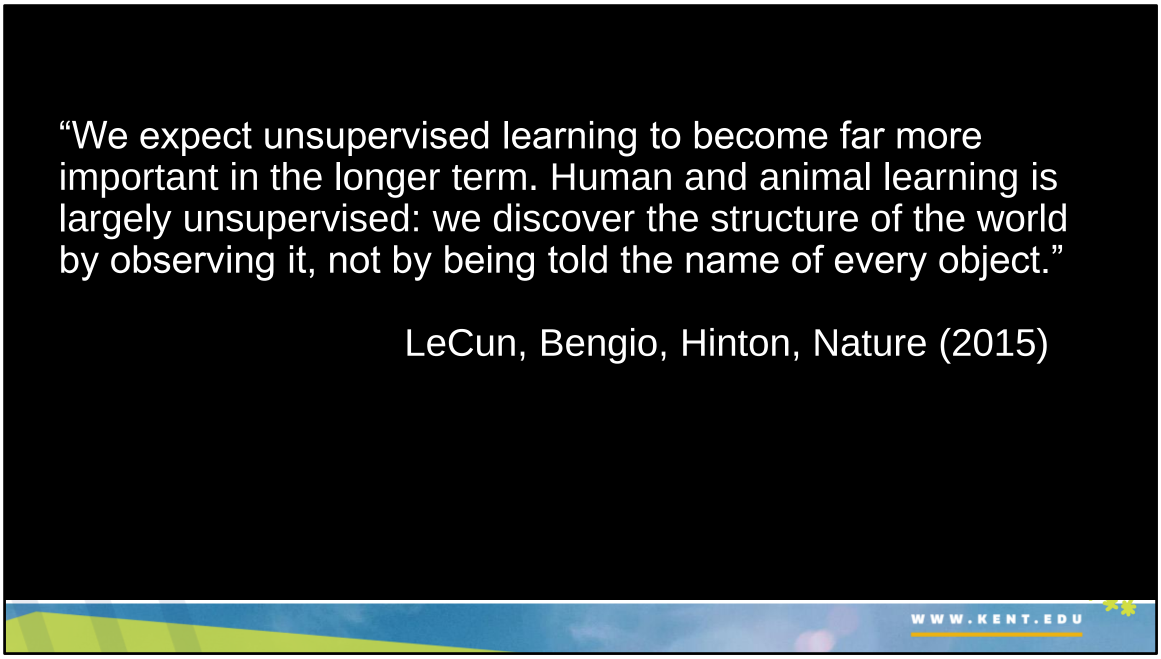
Reinforcement learning

- In reinforcement learning , the algorithm (agent) learns a policy of how to act given an observation of the world.
- Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm.
- Over the time, the algorithm (agent) learns to take a suitable action to maximize reward in a particular situation.



WWW.KENT.EDU

Reinforcement learning has lately been getting attention, especially after its success playing GO, and on multi-player games. Currently, reinforcement learning is mostly a research area and hasn't yet had significant practical successes beyond games. In time, however, we expect to see reinforcement learning take over an increasingly large range of real-world applications: self-driving cars, robotics, resource management, education, and so on. It's an idea whose time has come, or will come soon.



This concludes our lecture on types of ML models