LET'S START WITH WHAT IS MACHINE LEARNING ?

(AN INTRODUCTION TO MACHINE LEARNING)

SHAPEAI

- Machine Learning is the science (and art) of programming computers so they can learn from data.
- Here is a slightly more general definition:

> [Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.
>
> —Arthur Samuel, 1959

SPAM MAIL CLASSIFICATION USING MACHINE LEARNING ?

(SPAM OR HAM CLASSIFICATION)

SHAPEAI

Consider how you would write a spam filter using traditional programming techniques:

SHAPEAI

A spam filter based on Machine Learning techniques :

SHAPEAI

A spam filter based on Machine Learning techniques that automatically adapts to changes :

SHAPEAI

Finally, Machine Learning can help humans learn :

SHAPEAI

To summarize, Machine Learning is great for:

- Problems for which existing solutions require a lot of fine-tuning or long lists of rules: one Machine Learning algorithm can often simplify code and perform better than the traditional approach.
- Complex problems for which using a traditional approach yields no good solution: the best Machine Learning techniques can perhaps find a solution.
- Fluctuating environments: a Machine Learning system can adapt to new data.
- Getting insights about complex problems and large amounts of data.

APPLICATIONS OF MACHINE LEARNING ?

(HOW CAN WE USE MACHINE LEARNING)

SHAPEAI

- Analyzing images of products on a production line to automatically classify them:
- Detecting tumors in brain scans
- Automatically classifying news articles
- Automatically flagging offensive comments on discussion forums
- Summarizing long documents automatically
- Creating a chatbot or a personal assistant
- Forecasting your company’s revenue next year
- Making your app react to voice commands
- Detecting credit card fraud
- Segmenting clients based on their purchases so that you can design a different marketing strategy for each segment
- Recommending a product that a client may be interested in, based on past purchases

SUPERVISED/ UNSUPERVISED LEARNING

(TYPE OF SUPERVISION THEY GET DURING TRAINING)

SHAPEAI

There are four major categories:

- Supervised Learning
- Unsupervised Learning
- Semisupervised Learning, and
- Reinforcement Learning.

In this part we will go through all the types of learning trying to get a high level intution and understanding behind all the learning techniques

SHAPEAI

Supervised learning

In supervised learning, the training set you feed to the algorithm includes the desired solutions, called labels. We can the data being books which we use to teach the model(students).

SHAPEAI

SHAPEAI

SHAPEAI

Here are some of the most important supervised learning algorithms (covered in this course:

- k-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machines (SVMs)
- Decision Trees
- Random Forests, and
- Neural networks

SHAPEAI

Unsupervised learning

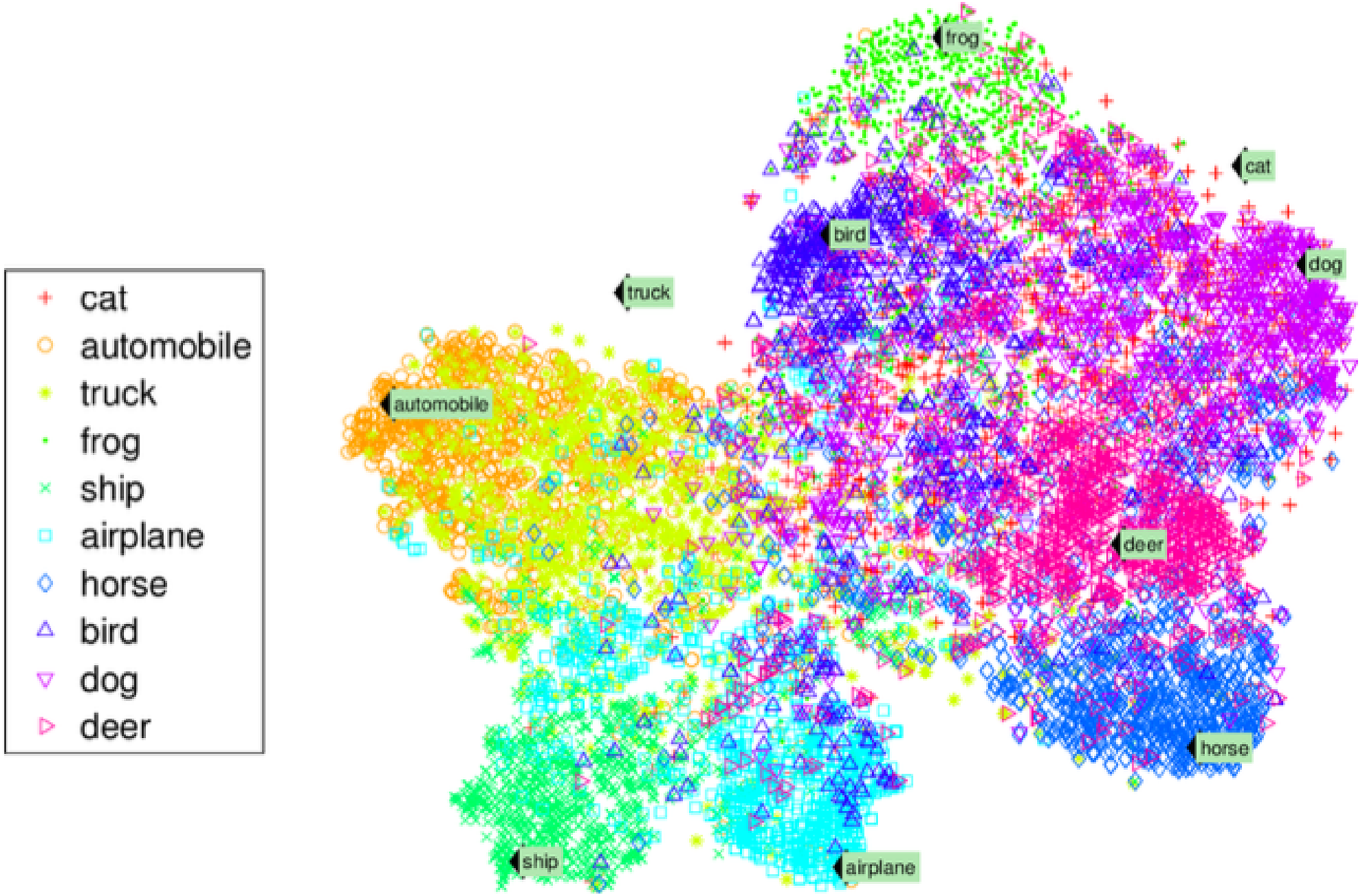
In unsupervised learning, as you might guess, the training data is unlabeled. The system tries to learn without a teacher.

SHAPEAI

Here are some of the most important unsupervised learning algorithms (covered in this course):

- Clustering
  - — K-Means
  - — DBSCAN
  - — Hierarchical Cluster Analysis (HCA)
- Anomaly detection and novelty detection
  - — One-class SVM
  - — Isolation Forest
- Visualization and dimensionality reduction
  - — Principal Component Analysis (PCA)
  - — Kernel PCA
  - — Locally Linear Embedding (LLE)
  - — t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Association rule learning
  - — Apriori
  - — Eclat

Visualization Algorithm



Example of a t-SNE visualization highlighting semantic clusters

SHAPEAI

Semisupervised learning

Since labeling data is usually time-consuming and costly, you will often have plenty of unlabeled instances, and few labeled instances. Some algorithms can deal with data that’s partially labeled. This is called semisupervised learning

SHAPEAI

Reinforcement learning

Reinforcement Learning is a very different beast. The learning system, called an agent in this context, can observe the environment, select and perform actions, and get rewards in return (or penalties in the form of negative rewards). It must then learn by itself what is the best strategy, called a policy, to get the most reward over time. A policy defines what action the agent should choose when it is in a given situation.

MAIN CHALLENGES OF MACHINE LEARNING

("BAD ALGORITHM" AND "BAD DATA")

SHAPEAI

Insufficient Training Data

It takes a lot of data for most Machine Learning algorithms to work properly. Even for very simple problems you typically need thousands of examples, and for complex problems such as image or speech recognition you may need millions of examples (unless you can reuse parts of an existing model called as transfer learning which will be discussed later).

Overfitting

Overfitting refers to a model that models the training data too well. Overfittinghappens when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.

Complex models such as deep neural networks can detect subtle patterns in the data, but if the training set is:

- Noisy
- Small
- Introduced sampling noise,

Then the model is likely to detect patterns in the noise itself.

SHAPEAI

SHAPEAI

Overfitting happens when the model is too complex relative to the amount and noisiness of the training data. Here are possible solutions:

- Simplify the model by selecting one with fewer parameters (e.g., a linear model rather than a high-degree polynomial model), by reducing the number of attributes in the training data, or by constraining the model.
- Gather more training data.
- Reduce the noise in the training data (e.g., fix data errors and remove outliers).
- Constraining a model to make it simpler and reduce the risk of overfitting is called regularization.