Unit 0
What is Machine Learning?

Prof. Phil Schniter

THE OHIO STATE UNIVERSITY

ECE 5307: Introduction to Machine Learning, Sp23

Outline

- What basic tasks do we want to perform?
- What is machine learning?
- How do we do machine learning?
- What will you learn in this course?

A few important tasks

The following tasks are key to many fields (e.g., engineering, science, medicine, business, etc.)

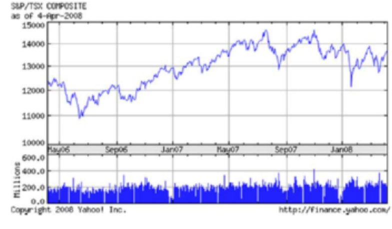
- Prediction (also known as Regression)
- Classification
- Categorization (also known as Clustering)

We'll describe these terms using a few examples...

Task 1: Prediction/Regression

Example:

- Say you want to use current & past data to predict a future stock price
- Could try to build a mathematical model of the stock market based on some economic theory, but that usually doesn't work very well



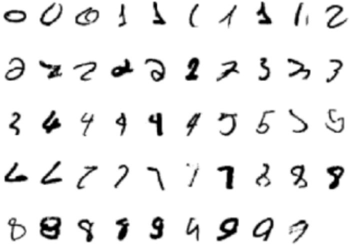
Stock trajectory

More generally, the goal of prediction is to estimate the value of one or more quantities given measurements of related data

Task 2: Classification

Example:

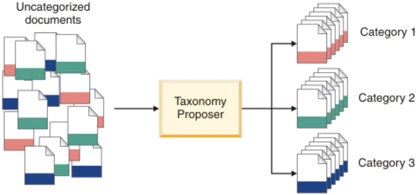
- Say you want to recognize a handwritten digit (in $\{0, 1, ..., 9\}$)
- Here again, it's difficult to come up with a mathematical model of the data
- But humans accomplish this task easily! How do we do it?



Handwritten digit examples

More generally, the goal of classification is to identify the category of a quantity given measurements of related data

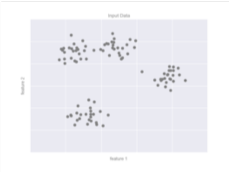
Task 3: Categorization/Clustering



Example:

- Say you are given a collection of text documents and you want to categorize them, but you have no preconceived idea of what the categories should be

More generally, the goal of clustering is to partition a given dataset into distinct groups, such that the data is "similar" within each group and "different" across groups



Outline

- What basic tasks do we want to perform?
- What is machine learning?
- How do we do machine learning?
- What will you learn in this course?

How we solve these tasks

- The **traditional approach** is to obtain a model from a scientific expert, and then apply that model to a given task
  - That works well in some applications (e.g., circuit design)
- But sometimes this traditional approach fails
  - A good scientific model may not exist (e.g., stock market)
  - A good scientific model may exist (e.g., acoustics) but may depend on parameters that are unknown (e.g., in noise cancellation)
- In **machine learning**, we build a model based on many examples of the data

Outline

- What basic tasks do we want to perform?
- What is machine learning?
- How do we do machine learning?
- What will you learn in this course?

The key ingredients

The key ingredients of machine learning are...

- a modeling function $f_{\boldsymbol{\theta}}(\cdot)$ that depends on many adjustable parameters $\boldsymbol{\theta} \in \mathbb{R}^d$
- a loss function $J(\cdot)$ that measures how well the model $f_{\boldsymbol{\theta}}(\cdot)$ is performing
- a training dataset on which to evaluate the loss (typically very large)
- an optimization algorithm that can find parameters $\boldsymbol{\theta}$ that yield a low loss on the training dataset

Example: machine learning for stock-market prediction

For example, in stock-market prediction...

- We want to predict a future stock price $y \in \mathbb{R}$ from currently available data $\boldsymbol{x} = [x_1, x_2, x_3, \ldots, x_d]^\mathsf{T} \in \mathbb{R}^d$
- The modeling function $f_{\boldsymbol{\theta}}(\cdot)$ should input $\boldsymbol{x}$ and output an accurate estimate of y
  - For $f_{\boldsymbol{\theta}}$, we might use a deep neural network. We'll study them later...
- The loss function $J(\cdot)$ quantifies accuracy
  - Using $\widehat{y} = f_{\boldsymbol{\theta}}(\boldsymbol{x})$ to denote our prediction of the true stock price y, we might choose $J(y, \widehat{y}) = (y - \widehat{y})^2$ or $J(y, \widehat{y}) = |y - \widehat{y}|$.
- The dataset $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$ would contain many past examples of the pair $(\boldsymbol{x}, y)$
- Our optimization algorithm aims to find the value of $\boldsymbol{\theta}$ that minimizes the loss on the training dataset:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \left\{ \frac{1}{n} \sum_{i=1}^n J\big(y_i, \underbrace{f_{\boldsymbol{\theta}}(\boldsymbol{x}_i)}_{\widehat{y}_i}\big) \right\}$$

- The result is a learned model $f_{\widehat{\boldsymbol{\theta}}}(\cdot)$ that aims to predict y from $\boldsymbol{x}$

Outline

- What basic tasks do we want to perform?
- What is machine learning?
- How do we do machine learning?
- What will you learn in this course?

What you'll learn in this course

In this course, you'll learn popular ways to...

- design the modeling function $f_{\boldsymbol{\theta}}(\cdot)$ (e.g., linear, polynomial, neural net)
- design the loss function $J(\cdot)$ (e.g., maximum likelihood, MAP, kernels)
- construct algorithms to optimize $\boldsymbol{\theta}$ (e.g., stochastic GD, Armijo)

for tasks like...

- prediction/regression
- classification, and
- categorization/clustering

You'll also learn many other important methods/concepts, e.g.,

- generalization, regularization, ensembling, PCA, NMF, GMMs, EM, ...

What you'll need for this course

The main goals are to learn...

- the concepts and mathematics behind machine learning
- how to implement machine learning using standard software tools, such as Python, NumPy, scikit-learn, and PyTorch

The assumed prerequisites are:

- undergraduate-level linear algebra
- undergraduate-level statistics
- scientific coding (although not necessarily in Python)

Homework #0 will review prerequisite knowledge of linear algebra, since it is so important to this course. A short Carmen quiz will follow

Questions?