Week 1:
Syllabus and Introduction to Machine Learning

CMPS 320 : Machine Learning

Welcome Message & Reminders

- Self Introductions and background
- Students Introductions
  - One interesting thing about yourself.
  - What's one big thing you hope to achieve out of this course?
- Review of course syllabus

Preparing for labs in Python

- Install Python on your own machine.
- Install Anaconda distribution on your system.
  - You can use many tools and applications in this course, without worrying about having to install and manage each separately.
  - The basic distribution includes Spyder as the IDE, IPython QtConsole, and Jupyter Notebook.
- There are several resources available on the internet for Python refresher:
  - Scientific Programming with Python
  - Swcarpentry: Plotting and Programming in Python

What Is Machine Learning?

- **Machine Learning** is the science (and art) of programming computers so they can learn from data.
- Machine learning can also be viewed as a set of computational tools for building statistical models.
- Definition by Tom Mitchell (1998): Machine Learning is the study of algorithms that
  - improve their performance P
  - at some task T
  - with experience E.

  A well-defined learning task is given by $< P, T, E >$

What Is Machine Learning? (Cont.)

Example:

- The spam filter is a Machine Learning program.
- Given examples of spam emails (e.g., flagged by users) and examples of regular (nonspam, also called "ham") emails, can learn to flag spam.
- The examples that the system uses to learn are called the **training set**.
- Each training example is called a training instance (or sample).
  - The task T, is to flag spam for new emails.
  - The experience E, is the training data, and
  - The performance measure P needs to be defined. For example,
    - ★ We can use the ratio of correctly classified emails.
    - ★ This performance measure is called accuracy, and it is often used in classification tasks.

Why Use Machine Learning?

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.

Machine Learning–Examples of Applications

- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images
- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant
- Prediction:
  - Company's revenue next year, based on many performance metrics
  - Future stock prices or currency exchange rates

Types of Machine Learning Systems

- Supervised (inductive) learning
  - Given: training data + desired outputs (labels)
- Unsupervised learning
  - Given: training data (without desired outputs)
- Semi-supervised learning
  - Given: training data + a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions

Supervised Learning

- The training set you feed to the algorithm includes the desired solutions, called **labels**.
- Typical supervised learning task are:
  - Classification
  - Regression
- Some supervised learning algorithms:
  - Linear Regression
  - Logistic Regression
  - Support Vector Machines (SVMs)
  - Decision Trees and Random Forests
  - Neural networks

Unsupervised Learning

- The training data is unlabeled.
  - The system tries to learn without a teacher.
- Some unsupervised learning algorithms
  - Clustering
    - ★ K-Means
    - ★ Hierarchical Cluster Analysis (HCA)
  - Visualization and dimensionality reduction
    - ★ Principal Component Analysis (PCA)
    - ★ Kernel PCA
    - ★ Locally Linear Embedding (LLE)
    - ★ t-Distributed Stochastic Neighbor Embedding (t-SNE)
  - Anomaly detection and novelty detection
    - ★ One-class SVM
    - ★ Isolation Forest
  - Association rule learning
    - ★ Apriori

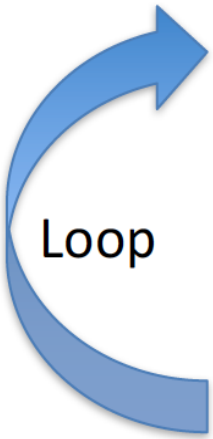
Semisupervised learning

- The training data is partially labeled.
  - Labeling data is usually time-consuming and costly.
- Most semisupervised learning algorithms are combinations of unsupervised and supervised algorithms.

Reinforcement Learning

- 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.
- Examples:
  - Many robots implement Reinforcement Learning algorithms to learn how to walk.
  - Game playing
  - Credit assignment problem

Machine Learning in Practice



- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn models
- Interpret results
- Consolidate and deploy discovered knowledge