Machine Learning Systems

M. R. Hasan
Readings

• Geron: chapter 1
Let’s discuss the **types of problems** we can solve with ML.

In other words, what **types of ML systems** we can build?

Let’s see…
We will discuss at least three different types of ML systems.

These are not mutually exclusive.

- Does it need human supervisor?
- Do we need to build a predictive model or just use the data to predict?
- Can it learn incrementally, or need the entire dataset?

Supervised, Unsupervised, Semi-supervised, Reinforcement

Model based, instance based

Batch and online learning
There are four major categories:

- Supervised learning, unsupervised learning, semisupervised learning, and Reinforcement Learning.

- **Supervised learning**: the training data for the algorithm includes the desired solutions (labels) created by humans.
Machine Learning: Human Supervision?

- There are **two types of supervised learning**.
- Classification
- Regression

**Classification**: Prediction is **categorical** (spam or ham).
Machine Learning: Human Supervision?

- Regression: Predict a numeric value.
- For example, price of a house based on features such as living area, no. of bedrooms, etc.
- Train the system with many examples of houses, including both their predictors and labels (i.e., their prices).
Some supervised algorithms that we will cover:

- *k*-Nearest Neighbors
- Linear Regression
- Logistic Regression
- Support Vector Machine (SVM)
- Artificial Neural Network
- Decision Tree and Random Forest
Unsupervised learning:
- The training data is unlabeled.
- The system tries to learn without a human supervisor.
- We will cover the following unsupervised algorithms:

**Clustering:**
- K-Means
- Expectation Maximization
- Gaussian Mixture Models

**Visualization & Dimensionality Reduction:**
- Principle Component Analysis
- Linear Discriminant Analysis
Semisupervised learning:

Some algorithms can deal with partially labeled training data:

Usually a lot of unlabeled data and a little bit of labeled data.

This is called semisupervised learning.

It’s a combination of supervised and unsupervised learning.
Machine Learning: Human Supervision?

• Semisupervised learning:
• Example: Google Photos.
• Once you upload all your family photos to the service, it automatically **recognizes that the same person A** shows up in photos 1, 5, and 11, while **another person B** shows up in photos 2, 5, and 7.
• This is the **unsupervised part** of the algorithm (clustering).
Machine Learning: Human Supervision?

- Semisupervised learning:
  - Now all the system needs is for you to tell it who these people are.
  - Just one label per person, and it is able to name everyone in every photo, which is useful for searching photos.
Machine Learning: Human Supervision?

- Reinforcement learning:
- Learning by **maximizing reward**.
Machine Learning: Human Supervision?

- Reinforcement learning:
- Example: DeepMind’s AlphaGo program.
- It learned its **winning policy by analyzing millions of games**, and then playing many games against itself.
Machine Learning

- Does it need human supervisor?

- Do we need to build a **predictive model** or just use the data to predict.

- Can it learn incrementally, or need the entire dataset?

Supervised, Unsupervised, Semi-supervised, Reinforcement

Model based, instance based

Batch and online learning
Machine Learning: Do We Need a Predictive Model?

• Most Machine Learning tasks are about making predictions.
• This means that given a number of training examples, the system needs to be able to generalize to examples it has never seen before.
• There are two main approaches to generalization:
  - Instance-based learning and
  - Model-based learning
Machine Learning: Do We Need a Predictive Model?

• Instance-Based Learning: We don’t need any model for prediction!
• Learn based on similarity.

For example, in the spam filter flag emails that are very similar to known spam emails.
Instance-Based Learning:

This requires a measure of similarity between two emails.

A (very basic) similarity measure between two emails could be to count the number of words they have in common.

The system would flag an email as spam if it has many words in common with a known spam email.
Machine Learning: Do We Need a Predictive Model?

- Model-Based Learning:
- Another way to generalize from a set of examples is to **build a model** of these examples.
- Then use that model to make predictions.

There are **two main approaches to generalization**:
- Instance-based learning and
- Model-based learning
Machine Learning

Does it need **human supervisor**?

Do we need to build a **predictive model** or just use the data to predict.

Can it **learn incrementally**, or need the entire dataset?

**Supervised**, **Unsupervised**, **Semi-supervised**, **Reinforcement**

**Model based**, **instance based**

**Batch and online learning**
Machine Learning: Can Learning be Incremental?

• In some problems, we need the **entire data set to train** the ML model.
• This is called **batch learning**.
• This will generally **take a lot of time** and computing resources, so it is typically done offline.
Machine Learning: Can Learning be Incremental?

• Alternative is to **learn in real-time** (as data arrives)!
Machine Learning: Can Learning be Incremental?

• In online learning, we train the system incrementally by feeding it data instances sequentially, either individually or by small groups called mini-batches.

• Each learning step is fast and cheap, so the system can learn about new data on the fly, as it arrives.
Machine Learning

- In summary, we will cover at least **three different types** of ML systems.

  **Does it need **human supervisor**?**

  **Do we need to build a **predictive model** or just use the data to predict?**

  **Can it **learn incrementally**, or need the entire dataset?**

  **Supervised, Unsupervised, Semi-supervised, Reinforcement**

  **Model based, instance based**

  **Batch and online learning**
Machine Learning

• There is yet another way of categorizing Machine Learning systems.
Machine Learning: Domingos’ Categorization

- **Five “tribes”** or schools of thoughts in Machine Learning.

<table>
<thead>
<tr>
<th>Tribe</th>
<th>Strength</th>
<th>Tools</th>
<th>Inspiration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbolists</td>
<td>Formal Systems</td>
<td>Decision Tree, Inductive Logic</td>
<td>Logic, Philosophy</td>
</tr>
<tr>
<td>Connectionists</td>
<td>Neural Networks</td>
<td>Backpropagation, Reinforcement Learning</td>
<td>Neuroscience</td>
</tr>
<tr>
<td>Bayesians</td>
<td>Probabilistic Inference</td>
<td>Hidden Markov Model (HMM), Graphical Models</td>
<td>Statistics</td>
</tr>
<tr>
<td>Evolutionaries</td>
<td>Structure Learning</td>
<td>Genetic Algorithms, Evolutionary Programming</td>
<td>Biology</td>
</tr>
<tr>
<td>Analogizers</td>
<td>Mapping to Novelty</td>
<td>K-NN, Support Vector Machine (SVM)</td>
<td>Psychology</td>
</tr>
</tbody>
</table>