Introduction

Intelligent Systems and Control

2019

Sepehr Maleki

University of Lincoln
School of Engineering

Machine Learning: What and Why?

“We” are generating more and more data.

- There are more than 1 trillion web pages.
- One hour of video is uploaded to YouTube per second.
- Tesco handles more than 6M transactions per day.

This deluge of data calls for automated methods of data analysis, which is what machine learning provides.

Machine Learning: What and Why?

"Data, is the oil of the digital era." – *The Economist, 6 May 2016*

Front Cover Cartoon by David Parkins.

(S. Maleki 2019)

Machine Learning– Definition

In particular, machine learning can be defined as a set of methods that can:

- Automatically detect patterns in data, and then use the uncovered patterns to predict future data.
- Perform decision making under uncertainty.

Here is a more concrete definition of machine learning:

"*Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in an autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.*"

(S. Maleki 2019)

Types of Machine Learning

1. **Supervised (predictive) learning** : Given a *labeled* set of input-output pairs $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$, the goal is to learn a mapping from inputs $\mathbf{x}$ to outputs y.

   $\mathcal{D}$ is called the *training set*, and N is the number of training examples.

   In the simplest setting, each training input $\mathbf{x}_i$ is a D-dimensional vector of numbers, representing a *feature*, say, the height and weight of a person.

   Similarly, the form of the output, y_i can in principle be anything, but most methods assume a *categorical* or *nominal* variable from some finite set, $y_i \in \{1, ..., C\}$ (such as male or female), or a real-valued scalar (such as income level).

(S. Maleki 2019)

Types of Machine Learning

When y_i is categorical, the problem is known as **classification** or pattern recognition, and when y_i is real-valued, the problem is known as **regression**.

N= 5

Features

Target

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	Price
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2

Figure: First 5 data of Boston Housing dataframe[1].

[1]This dataset contains information collected by the U.S Census Service concerning housing in the area of Boston Mass. It was obtained from the StatLib archive (http://lib.stat.cmu.edu/datasets/boston).

(S. Maleki 2019)

Types of Machine Learning

2. **Unsupervised (descriptive) learning**: Given a set of inputs, $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^{N}$, the goal is to find "interesting patterns" in the data.

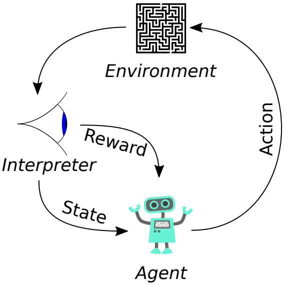
    - We do not know what kinds of patterns to look for.
    - There is no obvious error metric to use (unlike supervised learning, where we can compare our prediction of y for a given $\mathbf{x}$ to the observed value).

    Therefore, this is sometimes called **knowledge discovery** and is a much less well-defined problem.

(S. Maleki 2019)

Types of Machine Learning

3. **Reinforcement Learning**: Training of machine learning models (agents) to make a sequence of suitable decisions. The agent learns to achieve a goal in an uncertain, potentially complex environment.



(S. Maleki 2019)

Supervised Learning – Classification

Supervised learning is perhaps the most widely used type of machine learning algorithms.

In classification, we intend to learn a mapping from inputs $\mathbf{x}$ to outputs (targets) y, where $y \in \{1, \ldots, C\}$, with C being the number of classes.

- $C = 2$: This is called *binary classification* (in which case we often assume $y \in \{0, 1\}$).
- $C > 2$: This is called *multi-class classification*.

Supervised Learning – Classification

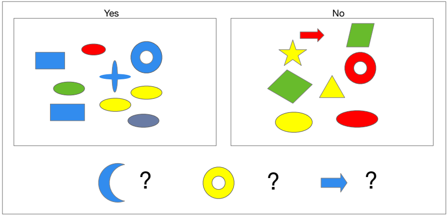
One way to define the classification problem is as *function approximation*.

We assume $y = f(\mathbf{x})$ for some unknown function f, and the goal of learning is to estimate the function f given a labeled training set.

When such f is learnt, we can make predictions using $\hat{y} = f(\mathbf{x}_{new})$ (hat symbol denotes an estimate) on novel inputs ($\mathbf{x}_{new}$) that we have not seen before (called *generalisation*).

Supervised Learning – Classification

Example: A simple toy classification problem.



D features

N Cases

Colour	Shape	Size (cm)	Label
Blue	Square	10	1
Red	Ellipse	2.4	1
Red	Ellipse	20.7	0

(S. Maleki 2019)

Supervised Learning – Classification

The need for probabilistic predictions!

To handle ambiguous cases, such as the yellow circle of the example, it is desirable to return a probability.

Denote the probability distribution over possible labels, given the input vector $\mathbf{x}$ and training set $\mathcal{D}$ by $p(y|\mathbf{x}, \mathcal{D})$ which represents a vector of length C.

In case of a binary classification, it is sufficient to return the single number $p(y = 1|\mathbf{x}, \mathcal{D})$, since:

$$p(y = 1|\mathbf{x}_{new}, \mathcal{D}) + p(y = 0|\mathbf{x}_{new}, \mathcal{D}) = 1 \ .$$

If the probability is further conditioned on some model, M, that we use to predict, we denote this by $p(y|\mathbf{x}_{new}, \mathcal{D}, M)$.

Supervised Learning – Classification

The need for probabilistic predictions!

Given a probabilistic output, we can always compute our "best guess" using:

$$\hat{y} = \hat{f}(\mathbf{x}) = \operatorname*{argmax}_{c=1}^{C} \; p(y = c | \mathbf{x}_{new}, \mathcal{D}) \ ,$$

which corresponds to the most probable class label, and is called the *mode* of the distribution $p(y|\mathbf{x}_{new}, \mathcal{D})$; it is also known as a MAP estimate (Maximum A-Posteriori).

Sometimes where $p(\hat{y}|\mathbf{x}_{new}, \mathcal{D})$ is far from 1.0 (i.e., we are not confident of our answer), it might be better to say "I don't know!" rather than returning an answer that we don't really trust.

Supervised Learning – Classification

Real-world application – Classifying flowers[1]

Versicolor (Type=1) Setosa (Type=0) Virginica (Type=2)

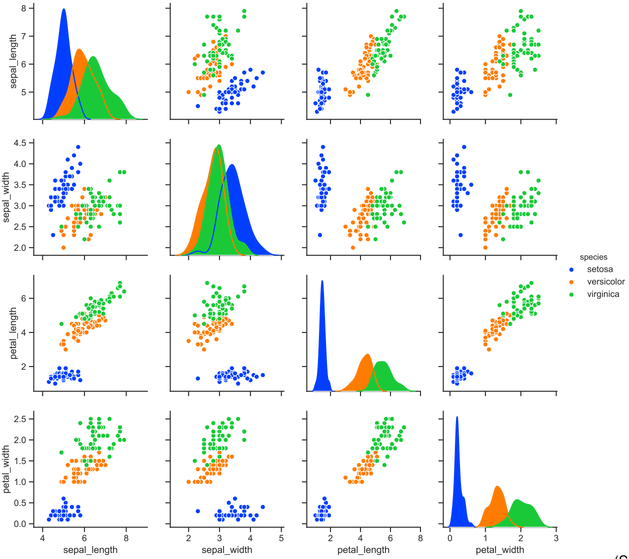
	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	Type
110	6.5	3.2	5.1	2.0	2
93	5.0	2.3	3.3	1.0	1
46	5.1	3.8	1.6	0.2	0
72	6.3	2.5	4.9	1.5	1
6	4.6	3.4	1.4	0.3	0

[1]Three types of iris flowers: Setosa, Versicolor and Virginica. Source: http://www.statlab.uni-heidelberg.de/data/iris/ .

(S. Maleki 2019)

Supervised Learning – Classification

Real-world application – Exploratory data analysis



(S. Maleki 2019)

Supervised Learning – Regression

Regression is just like classification except the response variable is continuous.

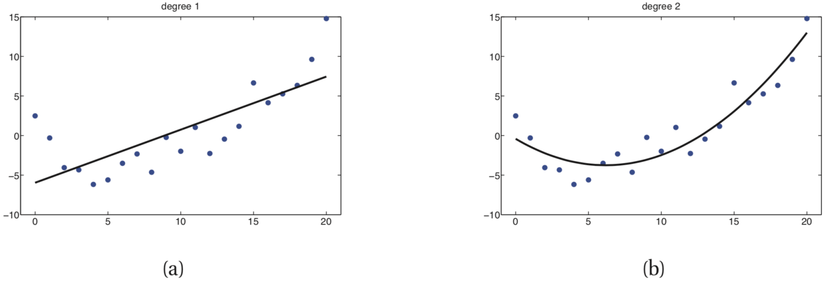


Figure: (a) Linear regression on some 1d data. (b) Same data with polynomial regression (degree 2).

(S. Maleki 2019)

Supervised Learning – Regression

Examples

Some examples of real-world regression problems include:

- Predict tomorrow's stock market price given current market conditions and other possible side information.
- Predict the age of a viewer watching a given video on YouTube.
- Predict the temperature at any location inside a building using weather data, time, door sensors, etc.
- Predict the location in 3d space of a robot arm end effector, given control signals (torques) sent to its various motors.
- Predict the amount of prostate specific antigen (PSA) in the body as a function of a number of different clinical measurements.

(S. Maleki 2019)