Data and Learning

Vivak Patel

Reading

1. Bousquet, Olivier, Stephane Boucheron, and Gabor Lugosi. "Introduction to statistical learning theory." Advanced lectures on machine learning. Springer, Berlin, Heidelberg, 2004. 169-207.
2. (Chapters 1, 2) Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. The elements of statistical learning. Vol. 1. No. 10. New York, NY, USA:: Springer series in statistics, 2001.

Outline

Between the lecture and homework, we will discuss and reflect on

1. Some philosophy about data and its uses. (Be critical of what can and cannot be done with data.)
2. Statistical Learning Theory. (Understand what statistical learning theory is attempting to do.)

Some Incomplete Philosophy

(1) The first definition that we will need is about **data**.

Some Incomplete Philosophy

(1) The first definition that we will need is about **data**.

(2) What is a sufficient condition for something to be data? (See Question 1 of Homework 1)

Some Incomplete Philosophy

I believe that the following are true:

Some Incomplete Philosophy

I believe that the following are true:

(1) Data is a collection of objects or information that is deemed (by someone) to be related.

Some Incomplete Philosophy

I believe that the following are true:

(1) Data is a collection of objects or information that is deemed (by someone) to be related.

(2) There are three primary uses of data.

- **(a)** Supply numerical evidence. (Statistics)
- **(b)** Make better or informed decisions. (Decision Theory)
- **(c)** Predict. (Learning)

Goal of this Course

Goal: To develop the (classical) mathematical theory of statistical learning.

Goal of this Course

Goal: To develop the (classical) mathematical theory of statistical learning.

Unfortunately, This theory, unfortunately, will be very narrow and unsatisfying.

1. We will focus on a specific type of statistical learning called **classification**.
2. We will develop the theory from very classical techniques to some limited modern techniques, but not all of them (e.g., deep neural networks).

Goal of this Course

Goal: To develop the (classical) mathematical theory of statistical learning.

Unfortunately, This theory, unfortunately, will be very narrow and unsatisfying.

1. We will focus on a specific type of statistical learning called **classification**.
2. We will develop the theory from very classical techniques to some limited modern techniques, but not all of them (e.g., deep neural networks).

However, while these classical techniques are obvious, they should be the first thing you reason about whenever you come across a prediction problem. They should *inform* your intuition.

Formalization

In order to specify what exactly we mean by "a mathematical theory of statistical learning," we need to precisely state several definitions.

Formalization

In order to specify what exactly we mean by "a mathematical theory of statistical learning," we need to precisely state several definitions.

(1) A **feature** is a vector x in $\mathbb{R}^d$. A **class** is a value in a finite-dimensional set $\{0, \ldots, M\}$ for some $M \in \mathbb{N}$. An **example** is a feature-class pair (x, y).

Formalization

In order to specify what exactly we mean by "a mathematical theory of statistical learning," we need to precisely state several definitions.

(1) A **feature** is a vector x in $\mathbb{R}^d$. A **class** is a value in a finite-dimensional set $\{0, \ldots, M\}$ for some $M \in \mathbb{N}$. An **example** is a feature-class pair (x, y).

(2) (Supervised) Learning is the process of generating a function $g : \mathbb{R}^d \to \{0, \ldots, M\}$. This function is called a **classifier**.

Formalization

(3) (Rough definition) **Statistical learning theory** is a collection of mathematical results based on *specific* assumptions and value judgments about learning from a collection of examples (i.e., data).

Formalization

(3) (Rough definition) **Statistical learning theory** is a collection of mathematical results based on *specific* assumptions and value judgments about learning from a collection of examples (i.e., data).

This definition should not be strange. In fact, the field of mathematical statistics has a rather similar definition. See Question 2 of Homework 1

Central Questions of SLT

Under a *precise* choice of assumptions and value statements:

Central Questions of SLT

Under a *precise* choice of assumptions and value statements:

(1) How can we define what a "best" classifier is?

Central Questions of SLT

Under a *precise* choice of assumptions and value statements:

(1) How can we define what a "best" classifier is?

(2) Does a best classifier exist?

Central Questions of SLT

Under a *precise* choice of assumptions and value statements:

(1) How can we define what a "best" classifier is?

(2) Does a best classifier exist?

(3) If it does, can we develop a procedure that supplies this best classifier? (This has nothing to do with whether we can actually compute the best classifier)

Central Questions of SLT

Under a *precise* choice of assumptions and value statements:

(1) How can we define what a "best" classifier is?

(2) Does a best classifier exist?

(3) If it does, can we develop a procedure that supplies this best classifier? (This has nothing to do with whether we can actually compute the best classifier)

Answering these questions will require that specify the assumptions and value judgments of SLT.