

# Mathematics for Machine Learning

Samuel Ireson  
samuel-a.ireson@db.com

Klejdí Sevdari  
klejdi-a.sevdari@db.com

Finlay Pillar  
finlay-pillar@db.com

Wing Lam Wong  
wing-lam-a-wong@db.com

October 30, 2024

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	What is learning? . . . . .	2
1.2	Structure of the course . . . . .	2
<b>2</b>	<b>Linear Algebra</b>	<b>4</b>
<b>3</b>	<b>Multivariable Calculus</b>	<b>5</b>
3.1	Introduction . . . . .	5
<b>4</b>	<b>Probability</b>	<b>6</b>
<b>5</b>	<b>Learning Problems</b>	<b>7</b>
5.1	Introduction . . . . .	7
5.2	A formal learning model . . . . .	7
5.2.1	Learning input . . . . .	7
5.2.2	Learning output . . . . .	7
5.2.3	A data generating model . . . . .	8
5.2.4	Measure of success . . . . .	8
<b>6</b>	<b>Optimisation Theory</b>	<b>9</b>
<b>7</b>	<b>Linear Regression</b>	<b>10</b>
<b>8</b>	<b>Logistic Regression</b>	<b>11</b>

# 1. INTRODUCTION

*We discuss motivations for studying the mathematics behind machine learning, give an outline of the content of the course, and mention the references which were used in putting it together.*

Machine learning is a term which is commonly heard and spoken, and one less commonly understood. This is in part due to the rapid growth in experimental machine learning techniques, and the lagging in mathematical explanations of these techniques. Another prominent reason is that the mathematical understanding of machine learning can, in enterprise settings, be considered as ‘less useful’ or ‘less relevant’. Some people may also perceive the inner workings of machine learning to be a ‘black box’, and a subject which they couldn’t understand.

We will focus on tackling the final obstruction. The perception of the value of understanding finds itself in the hands of others, but the understanding in question is firmly in our own. There are only a few mathematical concepts needed to gain a working understanding of machine learning, and we will cover these as they are needed.

We hope to present a self-contained course, covering the key topics in detail, gradually working towards popular models and algorithms which are seen in real-world machine learning applications. We will place a heavy focus on examples, ones which are hopefully relevant and familiar to the reader, rooting the work we are completing in the context of its application.

For those interested in reading more than we have time to cover, we recommend the work of Schwartz et al.[3], Bishop[1], and Faisal et al.[2], all of which were to a lesser or greater extent consulted in the writing of this course.

## 1.1 WHAT IS LEARNING?

Sometimes, it is those concepts of which we have an intuitive idea which are hardest to rigorously define. What does it mean to learn? What does it mean to have learned? What does it mean to understand? These are all relevant questions in the theory of machine learning, especially the first, and we will first consider them in the context of human learning.

In the loosest sense, *learning* is the process of converting past experiences into applicable knowledge. For humans, the input to this learning can come in any number of formats. For example, we could consider the knowledge obtained from reading a book. What exactly is the process occurring in the mind of the reader which allows them to retain, and importantly, re-apply this information. We note that the process is decidedly not memorisation. As a child who has memorised their times tables cannot claim the ability of multiplication, the reader who cannot call upon knowledge in unseen circumstances cannot claim to fully understand what they have read. This gives a good hint as to an important facet of learning systems; *generalisation*.

For machines, the process is not dissimilar. The input for a learning algorithm is data, and the output (the expertise) are some tuned parameters which allow the model produce outputs from unseen data in the future. There is more nuance to this, but as we move towards mathematical rigour, it will be useful to keep this in mind.

## 1.2 STRUCTURE OF THE COURSE

As previously mentioned, there are only few mathematical concepts which are absolutely necessary for understanding machine learning. An overview of the structure of this course is as follows,

- Linear Algebra

- Multivariable calculus
- Probability
- Optimisation
- Linear regression
- Logistic regression

Those familiar with the above topics may note that this course is structured in a ‘bottom-up’ fashion – we build on the theory as we move forwards. This is somewhat in contrast to other machine learning courses which take a ‘top-down’ approach in which the theory behind the application is uncovered as it is encountered. Both of these strategies have their strengths, and courses which are mathematically inclined typically take the same approach as us. It is however necessary to have patience with this approach. If you have previously only encountered machine learning in a practical, hands-on setting, you may struggle to understand how the early chapters of this course fit into the wider picture, but eventually all will be clear.

## 2. LINEAR ALGEBRA

## 3. MULTIVARIABLE CALCULUS

### 3.1 INTRODUCTION

## 4. PROBABILITY

Test

## 5. LEARNING PROBLEMS

*We now aim to combine the concepts which we have thus far encountered, and construct a formal definition for a learning problem. We will begin by defining the necessary inputs, and the desired outputs, before working towards the fundamental definition of PAC learnability, which will allow us to have some notion of the possibility of learning a certain problem.*

### 5.1 INTRODUCTION

We begin the chapter by laying out an example which we will reference and build around throughout. Suppose that you are a lender, aiming to determine if an arbitrary loanee will or will not default on the loan you give them. We phrase the problem deliberately in this manner, to emphasize the simplistic scenario in which we are working – the value of the loan is predetermined, and the goal is to provide a binary classification of loanees.

What information will be useful in this task? How will we determine if our solution is adequate? How will we provide a useful notation to describe the problem?

### 5.2 A FORMAL LEARNING MODEL

#### 5.2.1 LEARNING INPUT

The learner of a statistical learning problem has access to the following,

- Domain set: an arbitrary set which we typically label by  $\mathcal{X}$ . This set is nothing more than the objects to which we aim to assign labels. In the example outlined in the previous section, this set  $\mathcal{X}$  is the set of potential loanees.
- Label set: A set which we typically label by  $\mathcal{Y}$ . This is the set of possible labels, e.g., { default , no default }. Of course, it is more common to denote the possible labels simply by integers, and in our binary classification example, we will take  $\mathcal{Y} = \{\pm 1\}$ .
- Training data: a finite sequence of pairs  $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ , i.e., a sequence of labelled objects. In our case, this will be a sequence of past loaners together with information relating to their repayment. We will tend to denote the training data by  $\mathcal{S} = ((x_1, y_1), \dots, (x_m, y_m))$ , where  $m$  is the size of the training data.

REMARK. It is important to note that the training data is described as a sequence rather than a set. This is a rather nuanced point, and not one which we will dedicate much thought to. The reason for such pedanticity is that there exist learning algorithms which are dependent on the order of the training points, and there is no necessity for  $\mathcal{S}$  to be duplicate free.

It is however still common to refer to  $\mathcal{S}$  as the training set.

#### 5.2.2 LEARNING OUTPUT

- Prediction rule: the only output from a statistical learning problem is a function  $h : \mathcal{X} \rightarrow \mathcal{Y}$ . We also refer to the function as a hypothesis, and is a rule which the learner can use to label new elements of the domain space.



If we are considering an algorithm  $\mathcal{A}$ , then we will denote by  $\mathcal{A}(\mathcal{S})$  or  $h_{\mathcal{S}}$  the output hypothesis of the algorithm.

### 5.2.3 A DATA GENERATING MODEL

### 5.2.4 MEASURE OF SUCCESS

## 6. OPTIMISATION THEORY

## 7. LINEAR REGRESSION

## 8. LOGISTIC REGRESSION

# BIBLIOGRAPHY

- [1] Christopher Bishop. *Pattern Recognition and Machine Learning*. 2006.
- [2] Marc Peter Faisal Aldo; Deisenroth and Cheng Soon Ong. *Mathematics for Machine Learning*. 2020.
- [3] Shalev-Schwartz and Ben-David. *Understanding Machine Learning - from theory to algorithms*. 2014.