

# Notes on Machine Learning

Xi Tan (tan19@purdue.edu)

October 24, 2013



# Contents

<b>I</b>	<b>Introduction</b>	<b>5</b>
<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Some Distinctions . . . . .	7
1.1.1	Machine Learning v.s. Statistical Learning . . . . .	7
1.1.2	Parametric v.s. Non-parametric Models . . . . .	7
<b>2</b>	<b>A Brief History of Machine Learning</b>	<b>9</b>
<b>3</b>	<b>A Brief History of Machine Learning</b>	<b>11</b>
3.1	1950s-1960s: the Perceptron Model . . . . .	11
3.2	1970s-1980s: the Expert Systems . . . . .	11
3.3	mid 1980s: Multi-layer Neural Networks . . . . .	11
3.4	1995: SVM . . . . .	11
3.5	2000s: Logistic Regression Rediscovered . . . . .	11
<b>II</b>	<b>Supervised Learning</b>	<b>13</b>
<b>III</b>	<b>Unsupervised Learning</b>	<b>15</b>



## Part I

# Introduction



# Chapter 1

## Introduction

There are two main types of Machine Learning problems: the **predictive** or **supervised learning**, and the **descriptive** or **unsupervised learning** <sup>1</sup>.

For supervised learning, there are two subtypes: the **classification** problem, where the response variable is categorical (either ordinal or nominal); and the **regression** problem, where the response variable is numerical (either discrete or continuous).

For unsupervised learning, it is also called **density estimation** in Statistics literature. Essentially we want to build models of the form  $p(\mathbf{x}_i|\theta)$ , and supervised learning can be seen as to build models of the form  $p(y_i|\mathbf{x}_i, \theta)$ , which is a problem of conditional density estimation. <sup>2</sup>

### 1.1 Some Distinctions

#### 1.1.1 Machine Learning v.s. Statistical Learning

**Machine Learning** focuses more on high dimension low noise situations, and performance and efficiency are the main concerns. **Statistical Learning** focuses more on low dimension high noise situations, interpretation and statistical inference are the main concerns.

#### 1.1.2 Parametric v.s. Non-parametric Models

A parametric model has a fixed *finite* number of parameters, while the number of parameters in a non-parametric model grows with the amount of training data. A model with infinite number of parameters is usually non-parametric.

---

<sup>1</sup>It is also called **Knowledge Discovery** in the Data Mining literature

<sup>2</sup>We see from the formulation that, unsupervised learning usually involves multivariate probability models while supervised learning usually involves univariate probability models.





## Chapter 2

# A Brief History of Machine Learning

The perceptron model was invented in 1957, and it generated over optimistic view for AI during 1960s. After Marvin Minsky pointed out the limitation of this model in expressing complex functions, researchers stopped pursuing this model for the next decade.

In 1970s, the machine learning field was dormant, when expert systems became the mainstream approach in AI. The revival of machine learning came in mid-1980s, when the decision tree model was invented and distributed as software. It is also in mid 1980s multi-layer neural networks were invented, With enough hidden layers, a neural network can express any function, thus overcoming the limitation of perceptron. We see a revival of the neural network study.

Around 1995, SVM was proposed and have become quickly adopted.

After year 2000, Logistic regression was rediscovered and re-designed for large scale machine learning problems . In the ten years following 2003, logistic regression has attracted a lot of research work.

We discussed the development of 4 major machine learning methods. There are other method developed in parallel, but see declining use today in the machine field: Naive Bayes, Bayesian networks, and Maximum Entropy classifier (most used in natural language processing).

In addition to the individual methods, we have seen the invention of ensemble learning, where several classifiers are used together, and its wide adoption today.

[http://en.wikipedia.org/wiki/Timeline\\_of\\_artificial\\_intelligence](http://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence)



## Chapter 3

# A Brief History of Machine Learning

- 3.1 1950s-1960s: the Perceptron Model
- 3.2 1970s-1980s: the Expert Systems
- 3.3 mid 1980s: Multi-layer Neural Networks
- 3.4 1995: SVM
- 3.5 2000s: Logistic Regression Rediscovered



# **Part II**

## **Supervised Learning**



# Part III

## Unsupervised Learning

