

STANFORD UNIVERSITY

DOCTORAL THESIS

---

# Information, Prediction, and Supervised Learning

---

*Author:*

Charles ZHENG

*Supervisor:*

Dr. Trevor HASTIE and Dr.  
Jonathan TAYLOR

*A thesis submitted in fulfillment of the requirements  
for the degree of Doctor of Philosophy*

*in the*

Department of Statistics

March 7, 2017



## Declaration of Authorship

I, Charles ZHENG, declare that this thesis titled, “Information, Prediction, and Supervised Learning” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

---

Date:

---



Stanford University

# *Abstract*

Faculty Name  
Department of Statistics

Doctor of Philosophy

**Information, Prediction, and Supervised Learning**

by Charles ZHENG

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...



## *Acknowledgements*

The acknowledgments and the people to thank go here, don't forget to include your project advisor...





# Contents

<b>Declaration of Authorship</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Supervised learning . . . . .	1
1.2.1 General characterization of supervised learning . . . . .	1
1.3 Mutual information . . . . .	1
1.3.1 Definition and history . . . . .	1
1.3.2 Usage in neuroscience . . . . .	1
1.4 Generalizations of information . . . . .	1
1.4.1 Information axioms . . . . .	1
1.4.2 Information coefficients based on supervised learning . . . . .	1
<b>2 Randomized classification</b>	<b>3</b>
2.1 Motivation . . . . .	3
2.1.1 Facial recognition example . . . . .	3
2.2 Setup . . . . .	3
2.2.1 Sampling scheme . . . . .	3
2.2.2 Average accuracy . . . . .	3
2.3 Estimation of average accuracy . . . . .	3
2.3.1 Subsampling method . . . . .	3
2.3.2 Extrapolation . . . . .	3
2.4 Average Bayes accuracy . . . . .	3
2.4.1 Definitions . . . . .	3
2.4.2 Variance bound . . . . .	3
2.4.3 Inference of average Bayes accuracy . . . . .	3
<b>3 Extrapolating average accuracy</b>	<b>5</b>
3.1 Motivation . . . . .	5
3.1.1 Facial recognition example . . . . .	5
3.2 Assumptions . . . . .	5
3.3 Analysis of average risk . . . . .	5
3.4 Estimation . . . . .	5
3.5 Examples . . . . .	5
<b>4 Inference of mutual information</b>	<b>7</b>
4.1 Motivation . . . . .	7
4.1.1 Gene expression dataset example . . . . .	7
4.2 Identification loss . . . . .	7

4.3	Average Bayes accuracy and Mutual information . . . . .	7
4.4	Lower confidence bound . . . . .	7
4.5	Example . . . . .	7
<b>5</b>	<b>High-dimensional inference of mutual information</b>	<b>9</b>
5.1	Motivation . . . . .	9
5.1.1	Quantifying precision of decoding models . . . . .	9
5.1.2	Kay et al. example . . . . .	9
5.2	Setup . . . . .	9
5.3	Theory . . . . .	9
5.4	Estimator . . . . .	9
5.5	Examples . . . . .	9
<b>A</b>	<b>Frequently Asked Questions</b>	<b>11</b>
A.1	How do I change the colors of links? . . . . .	11

## **Chapter 1**

# **Introduction**

### **1.1 Introduction**

The study of complex systems.

### **1.2 Supervised learning**

The generalization error of the learner as a statistic.

#### **1.2.1 General characterization of supervised learning**

### **1.3 Mutual information**

#### **1.3.1 Definition and history**

#### **1.3.2 Usage in neuroscience**

### **1.4 Generalizations of information**

#### **1.4.1 Information axioms**

#### **1.4.2 Information coefficients based on supervised learning**



## Chapter 2

# Randomized classification

### 2.1 Motivation

#### 2.1.1 Facial recognition example

### 2.2 Setup

#### 2.2.1 Sampling scheme

#### 2.2.2 Average accuracy

### 2.3 Estimation of average accuracy

#### 2.3.1 Subsampling method

#### 2.3.2 Extrapolation

### 2.4 Average Bayes accuracy

#### 2.4.1 Definitions

#### 2.4.2 Variance bound

#### 2.4.3 Inference of average Bayes accuracy



## Chapter 3

# Extrapolating average accuracy

### 3.1 Motivation

#### 3.1.1 Facial recognition example

### 3.2 Assumptions

### 3.3 Analysis of average risk

### 3.4 Estimation

### 3.5 Examples





## Chapter 4

# Inference of mutual information

### 4.1 Motivation

#### 4.1.1 Gene expression dataset example

### 4.2 Identification loss

### 4.3 Average Bayes accuracy and Mutual information

### 4.4 Lower confidence bound

### 4.5 Example



## Chapter 5

# High-dimensional inference of mutual information

### 5.1 Motivation

#### 5.1.1 Quantifying precision of decoding models

#### 5.1.2 Kay et al. example

### 5.2 Setup

### 5.3 Theory

### 5.4 Estimator

### 5.5 Examples



## Appendix A

# Frequently Asked Questions

### A.1 How do I change the colors of links?

The color of links can be changed to your liking using:

```
\hypersetup{urlcolor=red}, or  
\hypersetup{citecolor=green}, or  
\hypersetup{allcolor=blue}.
```

If you want to completely hide the links, you can use:

```
\hypersetup{allcolors=.}, or even better:  
\hypersetup{hidelinks}.
```

If you want to have obvious links in the PDF but not the printed text, use:

```
\hypersetup{colorlinks=false}.
```