Dipartimento di Informatica, Bioingegneria,
Robotica ed Ingegneria dei Sistemi

Challenges in biomedical data science: data-driven solutions to clinical questions

by

Samuele Fiorini

Theses Series **DIBRIS-TH-2017-XX**

DIBRIS, Università di Genova
Via Opera Pia, 13 16145 Genova, Italy http://www.dibris.unige.it/

Università degli Studi di Genova

Dipartimento di Informatica, Bioingegneria,

Robotica ed Ingegneria dei Sistemi

Ph.D. Thesis in Computer Science and Systems Engineering
Computer Science Curriculum

Challenges in biomedical data science: data-driven solutions to clinical questions

by

Samuele Fiorini

September, 2017

Dottorato di Ricerca in Informatica ed Ingegneria dei Sistemi
Indirizzo Informatica
Dipartimento di Informatica, Bioingegneria, Robotica ed Ingegneria dei Sistemi
Università degli Studi di Genova

DIBRIS, Univ. di Genova
Via Opera Pia, 13
I-16145 Genova, Italy
http://www.dibris.unige.it/

Ph.D. Thesis in Computer Science and Systems Engineering
Computer Science Curriculum
(S.S.D. INF/01)

Submitted by Samuele Fiorini
DIBRIS, Univ. di Genova
. . . .

Date of submission: September 12, 2017

Title: Machine Learning 4 healthcare.

Advisor: Annalisa Barla
Dipartimento di Informatica, Bioingegneria, Robotica ed Ingegneria dei Sistemi
Università di Genova
. . .

Ext. Reviewers:
Lo Scopriremo
Lo Scopriremo
Lo Scopriremo

Abstract

Abstract

Contents

1 Introduction

Part I

2 Background

2.1 What is data science and why we should care

{

- **Data engineering**
- **Data exploration**
- **Machine learning and data understanding**
- **Data visualization**

}

2.2 Challenges in biomedical data science

2.3 From clinical questions to learning task

3 State of the art

3.1 Basic notation and definitions

3.2 Machine learning

3.2.1 Supervised learning

3.2.1.1 Regularization methods

3.2.1.2 Ensemble methods

3.2.1.3 Deep learning

3.2.2 Unsupervised learning

3.2.2.1 Manifold learning

3.2.2.2 Clustering

3.2.3 Model selection and evalutation

3.2.3.1 Model selection strategies

3.2.3.2 Feature selection stability

3.2.3.3 Performance metrics

Part II

4 ADENINE: a Data exploration tool

5 Model for biological age prediction [temp. title]

6 Temporal model for multiple sclerosis evolution

7 Temporal model for glucose predictions

8 Conclusion

Bibliography