Research on Machine Learning for Drug Discovery

CHEN, Wu

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy

in

Biomedical informatics

The Chinese University of Hong Kong

September 2020

Thesis Assessment Committee

Professor CHENG ** (Chair)

Professor ZHANG ** (Thesis Supervisor)

Professor CHEN ** (Committee Member)

Professor Li ** (External Examiner)

Abstract of thesis entitled:
Research on Machine Learning for Drug Discovery
Submitted by CHEN, Wu
for the degree of Doctor of Philosophy
at The Chinese University of Hong Kong in September 2020

The recent years have witnessed a rapid advance in machine learning (ML) algorithms and increasing availability of biomedical data. There is great opportunity to integrate the advances both fields to benefit healthcare.

近年来，机器学习算法得到了飞速发展，同时生物医学数据的获取也越来越容易。融合这两个领域的进步有利于医疗保健的发展。

Acknowledgement

I would like to thank my supervisor for offering me the opportunity to work with him.

Thanks everybody who helped me for their kindness!

Dedicated to those who risked their life to fight against COVID-19.

Contents

List of Figures

List of Tables

Symbols and Acronyms

In general, we denote a scalar by an italic lower case letter, a vector by a roman lower case bold letter, and a matrix by a roman upper case bold letter respectively, e.g., $a \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{M} \in \mathbb{R}^{p \times q}$, with any exceptions to be mentioned in the context case by case.

An identity matrix is written as $\mathbf{I}$. Specifically, an $n \times n$ identity matrix is written as $\mathbf{I}_n$. A zero matrix or vector is written as $\mathbf{0}$. Specifically, an $m \times n$ zero matrix is written as $\mathbf{0}_{m \times n}$.

Specialized symbols and major acronyms are defined as follows:

$\mu_0(\mathbf{x})$	a function for control group $f(\mathbf{x}, w = 0)$
$\mu_1(\mathbf{x})$	a function for treatment group $f(\mathbf{x}, w = 1)$
$p(\cdot)$	the probability density function (PDF)
$\Pr(\cdot)$	the probability value
$\mathbb{E}(\cdot)$	the expectation
$\mathbf{\Sigma}$	a covariance matrix
$\mathbf{N}(\mu, \mathbf{\Sigma})$	a normal distribution with mean μ and covariance $\mathbf{\Sigma}$
ε	a noise vector
$\mathbf{e}(\cdot)$	an error/residual function
$\mathbf{H}$	Hessian matrix
$\mathrm{tr}(\cdot)$	trace of a matrix
$\det(\cdot)$	determinant of a matrix

DNN	deep neuron network
GBM	gradient boosting machine
SVM	support vector machine

Chapter 1

Introduction

Machine learning (ML) is one of the fastest growing fields in science in the past decade.

□ **End of chapter.**

Chapter 2

Chapter title

□ **End of chapter.**

Chapter 3

Chapter title

□ **End of chapter.**

Chapter 4

Chapter title

□ **End of chapter.**

Chapter 5

Conclusions

Please conclude the thesis here.

□ **End of chapter.**

Chapter 6

Publication List

□ **End of chapter.**