

## DOCTORAL THESIS

### Stochastic gradient descent for pairwise learning: stability and optimization error

Shen, Wei

*Date of Award:*  
2019

[Link to publication](#)

#### General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

# Abstract

In this thesis, we study the stability and its trade-off with optimization error for stochastic gradient descent (SGD) algorithms in the pairwise learning setting. Pairwise learning refers to a learning task which involves a loss function depending on pairs of instances among which notable examples are bipartite ranking, metric learning, area under ROC curve (AUC) maximization and minimum error entropy (MEE) principle.

Our contribution is twofold. Firstly, we establish the stability results for SGD for pairwise learning in the convex, strongly convex and non-convex settings, from which generalization errors can be naturally derived. Moreover, we also give the stability results of buffer-based SGD and projected SGD. Secondly, we establish the trade-off between stability and optimization error of SGD algorithms for pairwise learning. This is achieved by lower-bounding the sum of stability and optimization error by the minimax statistical error over a prescribed class of pairwise loss functions. From this fundamental trade-off, we obtain lower bounds for the optimization error of SGD algorithms and the excess expected risk over a class of pairwise losses.

In addition, we illustrate our stability results by giving some specific examples and experiments of AUC maximization and MEE.

**Keywords:** Stability; Generalization; Optimization Error; Stochastic Gradient Descent; Pairwise Learning; Minimax Statistical Error.

# Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	vii
List of Tables	xii
Chapter 1 Introduction	1
1.1 Pairwise Learning . . . . .	1
1.1.1 AUC Maximization . . . . .	2
1.1.2 Metric Learning . . . . .	3
1.1.3 Minimum Error Entropy . . . . .	4
1.2 Stochastic Gradient Descent for Pairwise Learning . . . . .	5
1.2.1 Excess Expected Risk . . . . .	6
1.2.2 Generalization and Stability . . . . .	6
1.2.3 Optimization Error . . . . .	7
1.3 Main Contribution . . . . .	8
1.4 Related Work . . . . .	9
1.5 Outline of the Thesis . . . . .	10
Chapter 2 Background	12
2.1 Basic Concepts from Convex Optimization . . . . .	12

2.1.1	Elementary Definitions . . . . .	12
2.1.2	Co-coercivity of Convex Smooth Functions . . . . .	13
2.2	Le Cam’s Method for Minimax Lower Bound . . . . .	15
2.2.1	Minimax Risk in Statistics . . . . .	15
2.2.2	From Expected Risk to Hypotheses Testing . . . . .	16
2.2.3	Total Variation, Kullback-Leibler Divergence and Le Cam’s Method . . . . .	16
Chapter 3 Stability and Generalization of Stochastic Gradient Descent for Pairwise Learning . . . . . 21		
3.1	Preliminaries . . . . .	21
3.1.1	SGD for Pairwise Learning . . . . .	22
3.1.2	Algorithmic Stability and Its Relation with Generalization . . . . .	22
3.2	Warm-up: Some Technical Preparation . . . . .	24
3.3	Stability Analysis of SGD Algorithms . . . . .	32
3.3.1	Convex case . . . . .	32
3.3.2	Strongly convex case . . . . .	33
3.3.3	Non-convex case . . . . .	36
Chapter 4 Stability Analysis of Variants of Stochastic Gradient Descent for Pairwise Learning . . . . . 39		
4.1	Preliminaries . . . . .	39
4.1.1	SGD Algorithms using Finite Buffers . . . . .	40
4.1.2	Projected SGD Algorithms . . . . .	42
4.1.3	Projected SGD Algorithms using Finite Buffers . . . . .	45
4.2	Stability Analysis of SGD using Finite Buffers . . . . .	45
4.2.1	Some Technical Preparing . . . . .	45
4.2.2	Stability Analysis for Buffer-based SGD Under Random Selection Rule . . . . .	49
4.2.3	Stability Analysis for Buffer-based SGD Under Random Permutation Rule . . . . .	51
4.3	Stability Analysis for Projected SGD . . . . .	57
4.3.1	Same Results Inheriting from SGD . . . . .	57

4.3.2	Some New Results . . . . .	60
4.4	Stability Analysis for Projected SGD Using Finite Buffers . . . . .	65
4.5	Examples . . . . .	67
4.5.1	AUC Maximization . . . . .	68
4.5.2	Minimum Error Entropy Principle . . . . .	70
Chapter 5	Stability and Optimization Error Trade-off of Stochastic Gradient Descent for Pairwise Learning	72
5.1	Preliminaries . . . . .	72
5.1.1	Stability and Optimization Error Decomposition . . . . .	72
5.1.2	Loss and Distribution Independent Stability and Optimization Error	74
5.2	Trade-off between Stability and Optimization Error . . . . .	76
5.2.1	Minimax Lower Bounds . . . . .	76
5.2.2	Optimization Lower Bounds for SGD of Pairwise Learning . . . .	77
5.3	Proofs . . . . .	81
5.3.1	Proof of Theorem 5.2.1 . . . . .	81
5.3.2	Proof of Theorem 5.2.2 . . . . .	86
Chapter 6	Experiment Results	89
6.1	AUC Maximization . . . . .	89
6.2	Minimum Error Entropy Principle . . . . .	93
Chapter 7	Conclusion and Future Work	104
	Bibliography	106
	Appendix A: More Experiment Results	111
A.1	Experiments of SGD using Random Selection Rule for AUC Maximization on <code>fourclass</code> Data Set . . . . .	111
A.2	Experiments of SGD using Random Permutation Rule for AUC Maximiza- tion on More Data Sets . . . . .	112
A.3	Experiments of SGD using Random Selection Rule for MEE Principle . .	113
	Curriculum Vitae	133