Regularization for Regression and Ranking
回歸和排序中的正則化算法

Submitted to
Department of Mathematics
數學系
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
哲學博士學位

by

ZHAO Yulong
趙玉龍

March 2013
二零一三年三月
Abstract

Regularization is a method for learning and approximation which uses some additional information to avoid overfitting in statistics and machine learning. The information usually aims at improving the generalization ability by restrictions on regularity of potential functions. In this thesis, we mainly focus on the elastic net for regression and regularized least squares ranking algorithms.

The elastic net regularization is analyzed in two settings, according to their hypothesis spaces. One assumes a data independent hypothesis space composed by features independent of samples. Within this setting, significant contributions are made in several aspects. First, concentration estimates for sample error are presented by introducing $\ell^2$-empirical covering number and utilizing an iteration process. Second, a constructive approximation approach for estimating approximation error is presented. Third, the elastic-net learning with infinite features is studied and the role that the tuning parameter $\zeta$ plays is also discussed. Finally, our learning rate is shown to be faster compared with existing results. The other assumes a data dependent hypothesis space which is a subspace of a Reproducing Kernel Hilbert Space. Based on the capacity condition of the Reproducing Kernel Hilbert Space, a learning rate for elastic net is obtained by a stepping stone technique and an $\ell^2$-empirical covering number technique. The role of parameters is also discussed.

The regularized least squares ranking algorithm is analyzed in Reproducing Kernel Hilbert Spaces. By Hoeffding’s decomposition, a U-statistic could be decomposed
into an independent term and a degenerate U-statistic term. These two terms can be analyzed individually. The optimal learning rate is achieved.
Contents

Abstract i

Acknowledgement iii

1 Introduction 1

1.1 A Brief History of Learning Theory .......................... 1
1.2 A Framework of Supervised Learning .......................... 3
1.3 Regularization for Linear Regression ......................... 6
  1.3.1 Occam’s Razor ...................................... 7
  1.3.2 Ridge Regression .................................... 8
  1.3.3 Lasso ............................................. 10
  1.3.4 Elastic Net ......................................... 11
1.4 Regularization for Kernel Based Regression ................. 13
  1.4.1 Mercer Kernel ....................................... 13
  1.4.2 Reproducing Kernel Hilbert Space ..................... 15
  1.4.3 Kernel Ridge Regression ............................. 16
  1.4.4 Modified Kernel Projection Machine .................. 17
  1.4.5 Coefficient Based Kernel Regularization ............... 19
1.5 Thesis Outline and Contributions ............................ 21

2 Elastic Net Regularization in Data Independent Hypothesis Spaces 23
## Contents

2.1 Main Results on Elastic Net with Data Independent Hypothesis Spaces ........................................... 23  
2.2 Error Analysis and Estimation  ........................................................................................................... 28  
   2.2.1 Bounding Approximation Error ................................................................................................. 30  
   2.2.2 Concentration Estimates for Sample Error Involving $f_{\lambda}$ ............................................. 31  
   2.2.3 Concentration Estimates for Sample Error Involving $f_{\mu}$ ................................................... 32  
2.3 Improved Estimates via Iteration ....................................................................................................... 35  
2.4 Elastic Net Learning with Infinite Features ...................................................................................... 39  
2.5 Proofs of Error Bounds and Discussion  .......................................................................................... 42  
   2.5.1 Proofs of Error Bounds ............................................................................................................. 42  
   2.5.2 Toward Learning Rates and the Role that $\zeta$ Plays ............................................................... 43

3 Elastic Net Regularization in Data Dependent Hypothesis Spaces .................................................. 46  
   3.1 Introduction to Elastic Net with Data Dependent Hypothesis Spaces ........................................ 46  
   3.2 Main Result on Data Dependent Elastic Net ................................................................................. 48  
   3.3 Error Analysis ............................................................................................................................. 50  
      3.3.1 Error Decomposition ............................................................................................................. 51  
      3.3.2 Sample Error Estimation ..................................................................................................... 51  
      3.3.3 Hypothesis Error Estimation .............................................................................................. 54  
      3.3.4 Improved Hypothesis Error with Noise-free Distribution ............................................... 55  
   3.4 Convergence Rates of Data Dependent Elastic Net .................................................................... 57  
      3.4.1 Proofs of Theorem 8 and Theorem 9 .................................................................................. 57  
      3.4.2 Comparing Learning Rates ................................................................................................. 60

4 Regularized Least Squares Ranking Algorithm ............................................................................. 62  
   4.1 Introduction to Ranking .............................................................................................................. 62  
   4.2 Ranking Problem Formulation .................................................................................................... 64  
   4.3 Generalization Error Analysis .................................................................................................... 65  
      4.3.1 Sample Error Involving $f_{\lambda}$ ....................................................................................... 66
Contents

4.3.2 Sample Error Involving $f_{z,\lambda}$ ........................................ 68
4.3.3 Approximation Error ......................................................... 74
4.3.4 Generalization Bounds ..................................................... 74
4.4 Bounding Total Error by Iteration ......................................... 75
4.5 Comparisons with Related Work ........................................... 77
4.5.1 Comparison with Learning Rates for Regression ...................... 78
4.5.2 Comparison with Other Work for Ranking Algorithms ............ 78
4.6 Conclusions ................................................................. 79

5 Simulations and Future Work .................................................. 80

5.1 Simulation: Peptide Binding Affinity Prediction ......................... 80
5.1.1 Kernel Construction ......................................................... 81
5.1.2 Methods and Results ....................................................... 82
5.2 Future Work ................................................................. 83
5.2.1 Sparsity ................................................................. 83
5.2.2 Computation Related Topics ............................................. 84
5.2.3 Learning to Rank .......................................................... 84

Bibliography ................................................................. 87

Publications ............................................................... 98