# Contents

Declaration xi

Acknowledgments xiii

Abstract xv

1 Introduction 1

1.1 Learning Tasks ........................................... 2
  1.1.1 Supervised Learning .................................... 2
  1.1.2 Semi-Supervised and Transductive Learning .......... 3

1.2 Kernel Methods ............................................ 3

1.3 Greedy Algorithms ........................................ 4

1.4 Outline of the Thesis ..................................... 5

1.5 Bibliographical Notes ..................................... 7

I Deterministic Greedy Approaches 9

2 Knowledge-Based Kernel Principal Component Analysis 11

2.1 Kernel Principal Component Analysis ........................ 13
  2.1.1 Definition of Kernel Principal Components ........... 14
  2.1.2 Representer Theorem ................................... 16
  2.1.3 Derivation of Kernel Principal Components .......... 17

2.2 Semi-Supervised Learning .................................. 18
  2.2.1 Problem Setting ....................................... 18
  2.2.2 When Can Unlabeled Data Aid in Learning? .......... 19

2.3 Knowledge-Based Kernel Principal Component Analysis .. 20
  2.3.1 Definition of Knowledge-Based Kernel Principal Components .. 20
    2.3.1.1 Knowledge-Based Constraints ...................... 20
    2.3.1.2 Knowledge-Based Kernel Principal Components .... 21
  2.3.2 Representer Theorem .................................. 22
  2.3.3 Derivation of Knowledge-Based Kernel Principal Components .. 23

2.4 Optimization Problem ..................................... 25
## Contents

2.4.1 Eliminating Linear Constraint .................................. 26
2.4.2 Optimization of a Quadratic Form over a Hypersphere .... 27
2.4.3 Secular Equation .................................................. 29

2.5 Large Scale Approximations ........................................ 30
2.5.1 Iterative Optimization of a Quadratic Form over a Hypersphere . 30
  2.5.1.1 Iterative Computation of the Lagrange Multiplier ....... 30
  2.5.1.2 Conjugate Gradient Descent ................................ 31
2.5.2 Low-Rank Approximations ....................................... 34
  2.5.2.1 Nyström Method ............................................. 35
  2.5.2.2 Random Fourier Features ................................... 36

2.6 Interactive Data Visualization ...................................... 37
  2.6.1 Efficient Formulation of Interaction ......................... 37
  2.6.2 Rank-One Modification of a Diagonal Matrix ............... 38
    2.6.2.1 Secular Equation of a Rank-One Modification of a Diagonal Matrix ..................... 40
    2.6.2.2 Deflation .................................................. 41

2.7 Hyperparameter Optimization ...................................... 42

2.8 Discussion .......................................................... 45

2.9 Experiments ......................................................... 47

3 Greedy Feature Construction .......................................... 51
  3.1 Greedy Feature Construction ..................................... 52
    3.1.1 Overview ...................................................... 52
    3.1.2 Greedy Features ............................................. 53
    3.1.3 Convergence .................................................. 54
    3.1.4 Generalization Bound ....................................... 56
    3.1.5 Algorithm ...................................................... 61
  3.2 Learning with Fourier Features ................................... 63
    3.2.1 Fourier Features ............................................. 63
      3.2.1.1 Stationary Kernels ..................................... 63
      3.2.1.2 Random Kitchen Sinks ................................. 65
    3.2.2 Learning with Greedily Constructed Fourier Features .... 70
    3.2.3 À la Carte ..................................................... 75
  3.3 Experiments ........................................................ 77
    3.3.1 Baselines ..................................................... 78
    3.3.2 Results ........................................................ 79
  3.4 Discussion ........................................................ 80

II Randomized Greedy Approaches ................................. 85

4 Nyström Method with Kernel K-means++ Landmarks ............ 87
  4.1 Nyström Method .................................................... 88
    4.1.1 Nyström Method for Approximation of Integral Eigenfunctions .... 88
    4.1.2 Application of the Nyström Method to Low-Rank Approximation of Kernel Matrices ...... 90
    4.1.3 Alternative Derivation of the Nyström Method for Low-Rank Approximation of Kernel Matrices .................. 90
    4.1.4 Optimal Low-Rank Approximation of Kernel Matrix ........ 91
4.2 K-means Clustering .................................................. 93
  4.2.1 Optimization Problem ........................................... 93
  4.2.2 Relation to Optimal Low-Rank Approximation of Kernel Matrices . 94
4.3 Nyström Method with Kernel K-means++ Landmarks ................... 99
  4.3.1 K-means++ Sampling Scheme .................................... 99
  4.3.2 Kernel K-means++ Landmarks .................................... 101
  4.3.3 Theoretical Analysis ............................................. 102
4.4 Discussion ......................................................... 105
  4.4.1 Related Approaches .............................................. 105
  4.4.2 K-DPP Nyström Method ........................................... 106
  4.4.3 Instance Space K-means Centroids as Landmarks ................. 106
4.5 Experiments ......................................................... 108
4.6 Appendix: Additional Figures ....................................... 111
  4.6.1 Parkinsons ...................................................... 112
  4.6.2 Delta-Ailerons ................................................... 113
  4.6.3 Kinematics ...................................................... 114
  4.6.4 CPU-Activity .................................................... 115
  4.6.5 Bank .......................................................... 116
  4.6.6 Pumadyn ......................................................... 117
  4.6.7 Delta-Elevators .................................................. 118
  4.6.8 Ailerons ........................................................ 119
  4.6.9 Pole-Telecom ................................................... 120
  4.6.10 Elevators ....................................................... 121
  4.6.11 Cal-Housing .................................................... 122
  4.6.12 UJIL .......................................................... 123
  4.6.13 CT-Slice ....................................................... 124

5 Active Search for Computer-Aided Cyclic Discovery Processes 125
  5.1 Problem Setting and Algorithm ................................... 127
  5.2 Conditional Exponential Family Models ............................ 130
    5.2.1 Basic Notions .................................................. 130
    5.2.2 Relation to Conditional Max-Entropy Models .................. 131
    5.2.3 Relation to Gaussian Processes ................................ 135
  5.3 Markov Chains and the Metropolis–Hastings Algorithm ............ 137
    5.3.1 Markov Chains .................................................. 137
    5.3.2 The Metropolis–Hastings Algorithm ........................... 140
  5.4 Theoretical Analysis ............................................... 143
    5.4.1 Consistency .................................................... 143
    5.4.2 Mixing Time Analysis ......................................... 148
    5.4.3 Handling Large Importance Weights ............................ 150
  5.5 Adaptations to Exemplary Cyclic Discovery Processes .............. 150
    5.5.1 Discovery of Flavorful Cocktail Recipes ...................... 151
    5.5.2 Focused Drug Design Problem .................................. 153
    5.5.3 Synthetic Testbeds ............................................. 156
  5.6 Experiments ......................................................... 158
    5.6.1 Baselines ...................................................... 158
    5.6.2 Comparison against Baselines .................................. 159
    5.6.3 Drug Discovery ................................................. 161
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.7 Discussion</td>
<td>163</td>
</tr>
<tr>
<td>5.7.1 Machine Learning Perspective</td>
<td>163</td>
</tr>
<tr>
<td>5.7.2 De Novo Drug Design Perspective</td>
<td>165</td>
</tr>
<tr>
<td>5.8 Appendix</td>
<td>168</td>
</tr>
</tbody>
</table>

**List of Figures**

173

**List of Tables**

176

**Notation**

177

**Bibliography**

180