# Table des matières

Remerciements – Acknowledgements  
3

Au lecteur non-mathématicien  
5

Avertissement  
7

Foreword  
9

Chapitre 1. Introduction  
15
  1.1. Rééchantillonnage  
16
  1.1.1. L’heuristique d’Efron  
17
  1.1.2. Un peu d’histoire  
19
  1.1.3. Champs d’application  
20
  1.1.4. Avantages et limites  
20
  1.1.5. Résultats théoriques  
21
  1.2. Sélection de modèles  
23
    1.2.1. Cadre de la prédiction  
23
    1.2.2. Sélection de modèles  
28
  1.2.3. Pénalisation  
33
  1.2.4. Contributions de la thèse  
38
  1.3. Sélection de modèles par rééchantillonnage  
39
    1.3.1. Validation croisée  
39
    1.3.2. Pénalités par rééchantillonnage  
43
    1.3.3. Contributions de la thèse  
46
  1.4. Régions de confiance et tests par rééchantillonnage  
51
    1.4.1. Position du problème  
51
    1.4.2. Motivations pratiques  
56
    1.4.3. Méthodes par rééchantillonnage  
59
    1.4.4. Contributions de la thèse  
62

Notations  
65
  Conventions  
65
  Abbreviations  
65
  Mathematical notations (Chap. 2 to 9)  
65
  Histogram framework (Sect. 6.3, p. 152 and Sect. 6.8.1, p. 171)  
67
  Confidence regions and tests (Chap. 10)  
67

Chapter 2. Optimal model selection  
69
  2.1. Model selection for prediction  
70
    2.1.1. The prediction framework  
70
    2.1.2. Sharp oracle inequalities  
71
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2. A gap between theory and practice</td>
<td>74</td>
</tr>
<tr>
<td>2.2.1. Theory: hold-out, aggregation and local Rademacher complexities</td>
<td>74</td>
</tr>
<tr>
<td>2.2.2. Practice: V-fold cross-validation</td>
<td>77</td>
</tr>
<tr>
<td>2.3. Accurate calibration of penalties</td>
<td>78</td>
</tr>
<tr>
<td>2.3.1. A practical algorithm for calibration of penalties</td>
<td>79</td>
</tr>
<tr>
<td>2.3.2. The slope heuristics</td>
<td>79</td>
</tr>
<tr>
<td>2.3.3. Our contributions</td>
<td>80</td>
</tr>
<tr>
<td>2.4. Contributions on V-fold and other resampling procedures</td>
<td>81</td>
</tr>
<tr>
<td>2.4.1. Performances of V-fold cross-validation</td>
<td>81</td>
</tr>
<tr>
<td>2.4.2. V-fold and resampling penalties</td>
<td>82</td>
</tr>
<tr>
<td>Chapter 3. Slope heuristics</td>
<td>87</td>
</tr>
<tr>
<td>3.1. Introduction</td>
<td>87</td>
</tr>
<tr>
<td>3.2. Framework</td>
<td>88</td>
</tr>
<tr>
<td>3.2.1. Regression</td>
<td>88</td>
</tr>
<tr>
<td>3.2.2. Model selection</td>
<td>88</td>
</tr>
<tr>
<td>3.2.3. Histograms</td>
<td>89</td>
</tr>
<tr>
<td>3.3. Theoretical results</td>
<td>90</td>
</tr>
<tr>
<td>3.3.1. Optimal penalties</td>
<td>90</td>
</tr>
<tr>
<td>3.3.2. Minimal penalties</td>
<td>92</td>
</tr>
<tr>
<td>3.3.3. Comments</td>
<td>92</td>
</tr>
<tr>
<td>3.4. Practical use of slope heuristics: data-driven penalties</td>
<td>93</td>
</tr>
<tr>
<td>3.4.1. Computation of $\hat{K}_{\text{min}}$</td>
<td>93</td>
</tr>
<tr>
<td>3.4.2. Definition of $\hat{K}_{\text{min}}$</td>
<td>93</td>
</tr>
<tr>
<td>3.4.3. Shape of the penalty</td>
<td>94</td>
</tr>
<tr>
<td>3.4.4. Large number of models</td>
<td>95</td>
</tr>
<tr>
<td>3.5. Proofs</td>
<td>95</td>
</tr>
<tr>
<td>3.5.1. Proof of Thm. 3.1</td>
<td>95</td>
</tr>
<tr>
<td>3.5.2. Proof of Thm. 3.2</td>
<td>98</td>
</tr>
<tr>
<td>3.5.3. Proof of Prop. 3.1</td>
<td>100</td>
</tr>
<tr>
<td>3.5.4. Concentration inequalities used in the main proofs</td>
<td>101</td>
</tr>
<tr>
<td>3.5.5. Additional results needed</td>
<td>102</td>
</tr>
<tr>
<td>3.5.6. Proof of Prop. 3.3</td>
<td>102</td>
</tr>
<tr>
<td>3.5.7. Proof of Prop. 3.4</td>
<td>103</td>
</tr>
<tr>
<td>Chapter 4. Limitations of linear penalties</td>
<td>107</td>
</tr>
<tr>
<td>4.1. Introduction</td>
<td>107</td>
</tr>
<tr>
<td>4.2. Non-linearity of the ideal penalty in the histogram framework</td>
<td>108</td>
</tr>
<tr>
<td>4.3. Suboptimality of linear penalization</td>
<td>109</td>
</tr>
<tr>
<td>4.4. Simulation study</td>
<td>109</td>
</tr>
<tr>
<td>4.4.1. Framework of Prop. 4.1</td>
<td>110</td>
</tr>
<tr>
<td>4.4.2. Twelve more experiments</td>
<td>110</td>
</tr>
<tr>
<td>4.5. Proofs</td>
<td>117</td>
</tr>
<tr>
<td>Chapter 5. V-fold cross-validation</td>
<td>119</td>
</tr>
<tr>
<td>5.1. Introduction</td>
<td>119</td>
</tr>
<tr>
<td>5.2. Performance of V-fold cross-validation</td>
<td>121</td>
</tr>
<tr>
<td>5.2.1. General framework</td>
<td>121</td>
</tr>
</tbody>
</table>
6.8.2. Proof of Thm. 6.1 174
6.8.3. Proof of Thm. 6.1: alternative assumptions 174
6.8.4. Proof of Thm. 6.2 175
6.8.5. Proof of Lemma 6.4 177
6.8.6. Resampling constants 181
6.8.7. Concentration inequalities 183
6.8.8. Technical lemmas 186
6.8.9. Expectations of inverses 188

Chapter 7. The classification case 193
  7.1. Introduction 193
  7.2. Framework 195
  7.2.1. General framework 195
  7.2.2. Main assumptions 196
  7.2.3. Binary classification 197
  7.2.4. Bounded regression 197
  7.2.5. Resampling schemes 198
  7.3. Main results 198
  7.3.1. A recipe of oracle inequalities 198
  7.3.2. Concentration inequalities 199
  7.3.3. Program for further research 200
  7.4. Practical application 201
  7.4.1. Tuning parameters 202
  7.4.2. Comparison with other penalties 203
  7.5. Proofs 204
  7.5.1. Proof of Prop. 7.2 204
  7.5.2. Proof of Thm. 7.1 204
  7.5.3. $\varepsilon_{*,m}$ as a function of $n$ and $m$ 205

Chapter 8. Appendix on resampling penalties 207
  8.1. Uniqueness and existence of $\hat{s}_W$ 207
  8.1.1. Existence 207
  8.1.2. Uniqueness: histogram case 207
  8.1.3. Uniqueness: general case 208
  8.2. Resampling and structural constraints on the penalties 209
  8.3. Other assumption sets for oracle inequalities for RP 211
  8.3.1. Bounded case 211
  8.3.2. Unbounded case 211
  8.3.3. Sufficient condition for $(A\delta)$ 212
  8.4. Resampling Penalties with general weights 213
  8.4.1. Expectations 213
  8.4.2. Concentration inequality 214
  8.5. Useful concentration inequalities 215
  8.6. Moments, Exponential moments and Concentration 216
  8.6.1. Exponential moments $\Rightarrow$ concentration 216
  8.6.2. Moments $\Rightarrow$ concentration 217
  8.6.3. Moments $vs.$ Exponential moments 217
  8.6.4. Concentration $\Rightarrow$ moments 217
8.7. Expectation of inverses: symmetric case 218
Lower bound 219
Small values of $n$ 219
Upper bound 219
8.8. Concentration of inverses of multinomials: proofs 219
Lower deviations 220
Upper deviations 220
8.9. Moment inequalities for some U-statistics 223
Moments of $Z_1$ 223
Moments of $Z_2$ 223
8.10. Approximation properties of histograms 225
8.10.1. Results 225
8.10.2. Proofs 227

Chapter 9. On the constant in front of global penalties 233
9.1. Introduction 233
9.2. Lower bounds on $R_Z(\mathcal{F})$ 235
9.2.1. A factor 2 between theory and practice 235
9.2.2. Tight theoretical bounds 235
9.3. Upper bounds on $R_Z(\mathcal{F})$ 237
9.3.1. Two points classes 237
9.3.2. Independent classes 238
9.3.3. Rademacher complexities 239
9.4. Discussion 239
9.5. Proofs 240

Chapter 10. Resampling-based confidence regions and multiple tests 245
10.1. Introduction 246
10.1.1. Goals and motivations 246
10.1.2. Two approaches to our goal 247
10.1.3. Relation to previous work 247
10.1.4. Notations 248
10.2. Confidence region using concentration 249
10.2.1. Main result 249
10.2.2. Comparison in expectation 251
10.2.3. Concentration around the expectation 251
10.2.4. Resampling weight vectors 252
10.2.5. Practical computation of the thresholds 253
10.3. Confidence region using resampled quantiles 255
10.3.1. Main result 255
10.3.2. Practical computation of the resampled quantile 256
10.4. Application to multiple testing 257
10.4.1. Multiple testing and connection with confidence regions 258
10.4.2. Background on step-down procedures 259
10.4.3. Using our confidence regions to build step-down procedures 260
10.4.4. Uncentered quantile approach for two-sided testing 260
10.5. Simulations 262
10.5.1. Confidence balls 263
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5.2. Multiple testing</td>
<td>264</td>
</tr>
<tr>
<td>10.6. Discussion and concluding remarks</td>
<td>266</td>
</tr>
<tr>
<td>10.6.1. Confidence regions and tests</td>
<td>266</td>
</tr>
<tr>
<td>10.6.2. FWER versus FDR in multiple testing</td>
<td>267</td>
</tr>
<tr>
<td>10.6.3. About the variances of the coordinates</td>
<td>267</td>
</tr>
<tr>
<td>10.6.4. Conclusion</td>
<td>267</td>
</tr>
<tr>
<td>10.7. Proofs</td>
<td>268</td>
</tr>
<tr>
<td>10.7.1. Confidence regions using concentration</td>
<td>268</td>
</tr>
<tr>
<td>10.7.2. Quantiles</td>
<td>271</td>
</tr>
<tr>
<td>10.7.3. Multiple testing</td>
<td>274</td>
</tr>
<tr>
<td>10.7.4. Exchangeable resampling computations</td>
<td>274</td>
</tr>
<tr>
<td>10.7.5. Non-exchangeable weights</td>
<td>276</td>
</tr>
<tr>
<td>Chapter 11. Conclusions, open problems and prospects</td>
<td>281</td>
</tr>
<tr>
<td>11.1. Why should resampling be used?</td>
<td>281</td>
</tr>
<tr>
<td>11.1.1. A general purpose device</td>
<td>281</td>
</tr>
<tr>
<td>11.1.2. A naturally adaptive device</td>
<td>282</td>
</tr>
<tr>
<td>11.1.3. Choice of the resampling scheme</td>
<td>283</td>
</tr>
<tr>
<td>11.2. Advances in the non-asymptotic theory of resampling</td>
<td>284</td>
</tr>
<tr>
<td>11.2.1. Concentration results</td>
<td>284</td>
</tr>
<tr>
<td>11.2.2. Expectations</td>
<td>285</td>
</tr>
<tr>
<td>11.3. Optimal calibration of penalties</td>
<td>285</td>
</tr>
<tr>
<td>11.3.1. Flexibility of penalization</td>
<td>285</td>
</tr>
<tr>
<td>11.3.2. Calibration of resampling penalties with the slope heuristics</td>
<td>286</td>
</tr>
<tr>
<td>11.3.3. Need for overpenalization</td>
<td>287</td>
</tr>
<tr>
<td>11.3.4. Larger families of models</td>
<td>288</td>
</tr>
<tr>
<td>11.3.5. Global penalties</td>
<td>289</td>
</tr>
<tr>
<td>11.4. Confidence regions and multiple testing</td>
<td>289</td>
</tr>
<tr>
<td>11.4.1. Quantiles without additional term</td>
<td>289</td>
</tr>
<tr>
<td>11.4.2. Unknown noise-level</td>
<td>290</td>
</tr>
<tr>
<td>11.4.3. Non-gaussian or asymmetric data</td>
<td>290</td>
</tr>
</tbody>
</table>

Bibliographie                                                                 | 291  |