Contents

Preface v

Introduction vii

1. Inductive PAC-Bayesian learning 1

1.1. Basic inequality 2

1.2. Non local bounds 5

1.2.1. Unbiased empirical bounds 5

1.2.2. Optimizing explicitly the exponential parameter λ . . . 8

1.2.3. Non random bounds 9

1.2.4. Deviation bounds 11

1.3. Local bounds 14

1.3.1. Choice of the prior 14

1.3.2. Unbiased local empirical bounds 15

1.3.3. Non random local bounds 17

1.3.4. Local deviation bounds 18

1.3.5. Partially local bounds 22

1.3.6. Two step localization 27

1.4. Relative bounds 33

1.4.1. Basic inequalities 34

1.4.2. Non random bounds 37

1.4.3. Unbiased empirical bounds 40

1.4.4. Relative empirical deviation bounds 44

2. Comparing posterior distributions to Gibbs priors 51

2.1. Bounds relative to a Gibbs distribution 51

2.1.1. Comparing a posterior distribution with a Gibbs prior . . 52

2.1.2. The effective temperature of a posterior distribution . . 55

2.1.3. Analysis of an empirical bound for the effective temperature 56

2.1.4. Adaptation to parametric and margin assumptions . . . 61

2.1.5. Estimating the divergence of a posterior with respect to a Gibbs prior 67

2.2. Playing with two posterior and two local prior distributions 68

2.2.1. Comparing two posterior distributions 68

2.2.2. Elaborate uses of relative bounds between posteriors . . 70

2.2.3. Analysis of relative bounds 75

2.3. Two step localization . . . 89
2.3.1. Two step localization of bounds relative to a Gibbs prior . 89
2.3.2. Analysis of two step bounds relative to a Gibbs prior . . 96
2.3.3. Two step localization between posterior distributions . . 101

3. Transductive PAC-Bayesian learning . . . 111

3.1. Basic inequalities . . . 111
3.1.1. The transductive setting . . . 111
3.1.2. Absolute bound . . . 113
3.1.3. Relative bounds . . . 114

3.2. Vapnik bounds for transductive classification . . . 115
3.2.1. With a shadow sample of arbitrary size . . . 115
3.2.2. When the shadow sample has the same size as the training sample . . . 118
3.2.3. When moreover the distribution of the augmented sample is exchangeable . . . 119

3.3. Vapnik bounds for inductive classification . . . 121
3.3.1. Arbitrary shadow sample size . . . 121
3.3.2. A better minimization with respect to the exponential parameter . . . 123
3.3.3. Equal shadow and training sample sizes . . . 125
3.3.4. Improvement on the equal sample size bound in the i.i.d. case 125

3.4. Gaussian approximation in Vapnik bounds . . . 127
3.4.1. Gaussian upper bounds of variance terms . . . 127
3.4.2. Arbitrary shadow sample size . . . 128
3.4.3. Equal sample sizes in the i.i.d. case . . . 128

4. Support Vector Machines . . . 131

4.1. How to build them . . . 131
4.1.1. The canonical hyperplane . . . 131
4.1.2. Computation of the canonical hyperplane . . . 132
4.1.3. Support vectors . . . 134
4.1.4. The non-separable case . . . 134
4.1.5. Support Vector Machines . . . 138
4.1.6. Building kernels . . . 140

4.2. Bounds for Support Vector Machines . . . 142
4.2.1. Compression scheme bounds . . . 142
4.2.2. The Vapnik–Cervonenkis dimension of a family of subsets 143
4.2.3. Vapnik–Cervonenkis dimension of linear rules with margin 145
4.2.4. Application to Support Vector Machines . . . 148
4.2.5. Inductive margin bounds . . . 149

Appendix: Classification by thresholding . . . 155

5.1. Description of the model . . . 155

5.2. Computation of inductive bounds . . . 156

5.3. Transductive bounds . . . 158

Bibliography . . . 161