The handle http://hdl.handle.net/1887/33204 holds various files of this Leiden University dissertation

Author: Ommen, Thijs van
Title: Better predictions when models are wrong or underspecified
Issue Date: 2015-06-10
Bibliography


A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information

S. Ghosal, J. Ghosh, and A. van der Vaart. Convergence rates of posterior dis-

R. D. Gill. The Monty Hall problem is not a probability puzzle (It’s a challenge

R. D. Gill and P. D. Grünwald. An algorithmic and a geometric characterization

R. D. Gill, M. J. van der Laan, and J. M. Robins. Coarsening at random: Char-
acterizations, conjectures, counter-examples. In D. Y. Lin and T. R. Fleming,

A. V. Gnedin. The Monty Hall problem in the game theory class. *arXiv preprint

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, predictions, and
2007.

A. V. Goldberg and R. E. Tarjan. A new approach to the maximum-flow prob-

G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a
method for choosing a good ridge parameter. *Technometrics*, 21:215–223,
1979.


Available as ILLC Dissertation Series 1998-03; see www.grunwald.nl.

P. D. Grünwald. Viewing all models as “probabilistic”. In *Proceedings of the
Twelfth ACM Conference on Computational Learning Theory (COLT’ 99)*, pages


P. D. Grünwald. That simple device already used by Gauss. In P. D. Grünwald,
of Jorma Rissanen on the Occasion of his 75th Birthday*, pages 293–304. Tampere


Index

2-connected component, 164

affine transformation, 140, 155
AIC, 5, 9, 18–24, 26, 30, 107
Akaike weights, 25, 27

basis exchange, 158
Bayesian prior belief, see prior distribution
Bayesian Lasso, 42, 91
Bayesian predictive distribution, see predictive distribution
BIC, 18, 31, 32, 73, 107
BMA, 7, 19, 31, 32, 39
BMS, 18, 31, 32, 39, 58
BPIC, 18, 33
Bregman score, 155

capacitated Edmonds-Gallai decomposition, 195
capacity, 189, 201
CAR, 114, 162, 169–171
weak and strong, 115, 139
Cesàro-averaged posterior, 55, 61
circuit, electrical, 186–194
comasening mechanism, 114, 115, 119
colouring, 162
homogeneous, 162–167
induced, 162–167
componentwise rescaling, 193
maximum, 194
concavity (of entropy), 125
strict, 139
concentration, see posterior concentration

conditioning, 12, 117, 160
Jeffrey, 141
naive, 117, 141, 152, 171
standard, 114, 117
conductance, 187
connected game, 152
consistency, 40, 59–61, 89, 90
contestant, 116
continuity, 125
convex optimization, 124, 181
convexity (of a model), 72, 75
covariate, 17, 20, 40, 108
covariate shift, 24, 33
cross-validation, 18
10-fold, 107
generalized, 28, 32, 107
leave-one-out, 28, 91, 107, 108
current, electrical, 187–194
data compression, 2, 85, 86, 88
decomposition, 152
substitution, 153
demand, 190
design matrix, 20, 49
design vector, 20
DIC, 33
differentiability (of entropy), 136
diode, 186
discard
a message, 129, 168, 169
a message-outcome pair, 130, 156
dominating hyperplane, 126, 131
electrical circuit, 186–194
empirical Bayes, 42, 66, 100
Index

entropy, generalized, 118, 121
equalizer strategy, 133, 139
error
  extra-sample, 20, 26
generalization, 19
in-sample, 20
  squared, see loss function; risk
exchange (of outcomes), 156
exchange-connected game, 162, 163
extra-sample AIC, 5, 9, 23, 29

f-component, 191
FAIC, 6, 10, 25, 29
FIC, 28, 32
Fisher information, 21
flow, 190
  lexicographically maximum, 196
  maximum, 192, 198
  proportional, 192, 197
  under c, 192
flow conservation law, 190
focus parameter, 28

game, 118
  connected, 152
  exchange-connected, 162, 163
graph, 157–158, 160–161, 167
  bipartite, 133, 199
  path, 182
matroid, 158–161, 164, 167
maximin, 121, 132
minimax, 121, 132
negation, 159
online, 141
partition matroid, 168–169
sunflower, 169
uniform, 162
  zero-sum, 121, 122, 160
Gaussian distribution, 4, 9, 23
GCV, 28, 32, 107
generalized entropy, 118, 121
Gibbs error, 51
goat, 11, 114
graph, see game, graph
ground set, 158
hetero-/homoskedasticity, 42, 85
horse race, 140
hypercompression, 75–78, 83
hypergraph, 152
hyperplane, 126, 131
  dominating, 126, 131
  realizable, 131
  supporting, 126
    minimal, 131, 134, 136
I-component, 188
in-liker, 42
in-model loss, 53
inconsistency, 19, 25, 30, 60, 90
independent set, 159
input variable, 3, 17, 20, 40, 108
inverse gamma distribution, 50
Kelly gambling, 140
Kirchhoff’s current law, 188
KL divergence, 9–10, 19, 40, 45, 74
KL-optimal, 43, 45, 59, 73
KT-vector, 127, 133, 134, 136, 156
learning rate, 8, 40
  critical, 83
loss function, 8, 13, 47, 86, 90, 116,
  118, 122, 160
  0-1, 13, 116
    hard, 124, 132
    randomized, 124, 132
Brier, 116, 122, 129, 130, 161
for point predictions, 8, 32, 86
inherently asymmetric, 155
local, 138, 158
logarithmic, 8–10, 13, 45, 48,
  73, 76, 86, 108, 116, 122,
  127, 129, 138, 161
  skewed, 155
matrix, 155, 157
proper, see proper
squared error, 8–10, 20, 28, 32,
  40, 46, 108
symmetric, 154
  fully, 155
  w.r.t. exchanges, 156
loss invariance, 160–161
MAP model, 58–61
marginal distribution, 115, 118
matching, fractional, 200
matroid, 158, see also game, matroid
  basis packing, 201
  contraction, 202
  cycle, 159, 203
  independence testing oracle, 203
  partition, 168–169
  rank function, 201
  restriction, 202
  uniform, 159
maximin game, 121, 132
message, 12, 116
  dominated, 152
message structure, 14, 118,
  149–150, 161–164, 168–169
  dual, 170
minimax, 116, see also worst-case optimal
game, 121, 132
  optimal decision making, 118
minimax theorem, 122
minimum cut, 195
misspecification, 2, 22, 40, 73, 75,
  86–91
mix-loss, 81, 93
mixability gap, 79–84
Möbius strip, 186
model, 1, 2, 45
  conditional, 20
  linear, 4, 20, 46
model averaging, 5
  Bayesian, 7, 19, 31, 32, 39
  with Akaike weights, 25, 27
model selection, 4, 18
  Bayesian, 18, 31, 32, 39, 58
  extra-sample, 24
  focused, 6, 25
  module, 153, 164
Monty Hall problem, 11–12, 114,
  127, 158, 160
Nash equilibrium, 14, 122, 132, 136,
  155
network, 190
Ohm’s law, 187
optimal, see KL-optimal; worst-case optimal
outcome, 12, 113
  merging, 154
outcome space, 118
outlier, 42, 85
output variable, 3, 17, 20
overconfidence, 59, 81
partition, 152, 170
partition matroid, 168–169
pay-off, 140
possibly unsaturated, 201, 213, 215
posterior concentration, 51, 72, 74,
  79, 80, 94
  minimax optimal rate, 55
posterior distribution, 6, 115
  Cesàro-averaged, 55, 61
  generalized, 8, 10, 40, 47, 48
posterior-randomized loss, 51, 76,
  83
prediction, 1–3, 5–7, 13, 18, 116, 119
  extra-sample, 24
  worst-case optimal, 117
predictive distribution, 7, 32, 73, 74
  flattened generalized, 81
  generalized, 48
  variance, 78
prequential, 50, 106
prior distribution, 6, 99–106, 115
  Jeffreys’, 28, 56
  worst-case optimal, 117
prior predictive check, 79, 81
probability updating, 12, 115
  game, 118
  proper, 119, 120, 135
  strictly, 119
pseudo-truth, see KL-optimal
quizmaster, 116
randomized loss, see posterior-randomized loss; loss function (0-1; matrix)
RCAR, 139–141, 157, 159–161
   strong, 139
RCAR vector, 139, 164–166
realizable hyperplane, 131
redundancy, 73
regression function, true, 46
reliability, 47, 59, 60, 67, 69
resistor, 187
ridge regression, 7, 61–70, 100, 102
   (empirical) Bayesian, 42, 66, 68, 70
risk
   logarithmic, 45, 73–78
   squared, 28, 46–47, 58–60, 67, 69, 73, 76
SafeBayes, 8, 10, 42, 60
   algorithm, 52, 93
   I-log-, 54, 69, 70
   I-square-, 53, 70, 99–103, 108
   R-log-, 11, 54, 69, 70
   R-square-, 53, 70, 99–102
saturated, 193
scoring rule, see loss function
self-confidence ratio, 59
SIC, 28, 32
sink, 190
small sample correction, 24
source, 190
spike-and-slab data, 27, 57, 111
stable set, 133
strategy, 119
   contestant, 120
degenerate, 156
   equalizer, 133, 139
nondegenerate, 156
   quizmaster, 120
   RCAR, see RCAR
worst-case optimal, see
   worst-case optimal
strongly polynomial, 181
sunflower, 169
supergradient, 126, 131, 136
supersink/-source, 190
supervised learning, 17
supply, 190
supply proportion, 191
supporting hyperplane, see
   hyperplane, supporting
taut string, 183
   algorithm, 184
test data, 4–6, 17
   TIC, 21
training data, 3, 17
underspecification, 2, 5, 13, 141
   uniform game, 162
   uniform multicover, 170
updating
   minimum relative entropy, 141
   probability, see probability
   updating
utilization, 191
variance
   fixed, 20, 23, 99
   unknown, 24
voltage, 186, 191
voltage drop, 187, 191
WAIC, 18, 33
worst-case optimal, 13–14, 116, 160
   for the contestant, 121,
   130–136, 138, 140, 150, 160
   for the quizmaster, 14, 121,
   124–130, 136, 138–140,
   157, 159–161
wrong, 2, 84, see also
   misspecification
XAIC, 5, 9, 23, 29
zero-sum game, 13, 121, 122, 160