EM /
Self-Learning /
Assumptions

Marco Loog /
Jesse H. Krijthe

KU /
TU
Delft

- Somewhat different take than Lars Kai’s [very refreshing] view
  - General data [not images]
  - Blind for underlying data distribution
  - Aim at automation / “it should always work”
  - “Pessimistic”? Realistic?

- Main message : think about the assumption that you [implicitly or explicitly] make when using SSL techhniques!

- How to learn the parameters in models [for instance] considered by Castelli & Cover?

- First the model
  - Likelihood formulation
  - Integrating out for single observation
  - Then likelihood of full set of observations

- Classical [and de facto] approach is through expectation maximization [EM]
  - Alternate maximizing likelihood parameters and determining expected labels

- What if we do not have underlying statistical model?
  - E.g. what to replace the EM steps with?
  - Referred to as self-learning, self-training, pseudo-labelling, Yarowsky algorithm
  - Appealing idea, so has been reinvented a dozen times
  - [What happens if we do this with our stats model?]

- So what do we do if we self-learn SVMs?
  - Bridge to low-density separation by considering hinge loss
  - Sort of minmin optimization?
  - Leads to transductive SVMs and one way to train it

- Think about how many SSL method actually fit self-learning view
  - Co-training, label propagation, EM, more?

- Back to assumption
  - Accepted idea : we have to relate $P(x)$ to $P(y|x)$
  - Low-density separation is one way
  - Manifold and cluster assumption are also often used
  - [No-go theorem?]

- Can do great things if assumptions are correct
  - “Theory” : if assumptions on data hold, SSL works
  - Example of yesterday : works well...
  - Can also improve *supervised* methods if assumption is correct!
  - So, what happens if assumption is wrong?
  - When would EM, or likelihood in general, go wrong?
  - Can discounting help? In what sense?

- Qs?

A completely biased list of references

1. Abney, “Understanding the Yarowsky algorithm,” Computational Linguistics, 2004
2. Ben-David, Lu, Pal, “Does unlabeled data provably help?”, COLT, 2008
3. Bennett, Demiriz, “Semi-supervised support vector machines”, NIPS, 1999
4. Chapelle, Schölkopf, Zien, “Semi-supervised learning”, MIT press, 2006
5. Cohen, Cozman, Sebe, Cirelo, Huang, “Semisupervised learning of classifiers,” IEEE TPAMI, 2004
6. Cozman, Cohen, “Risks of semi-supervised learning,” in Semi-Supervised Learning. MIT Press, 2006
7. Grandvalet, Bengio, “Semi-supervised learning by entropy minimization”, NIPS, 2005
8. Haffari, Sarkar, “Analysis of semi-supervised learning with the Yarowsky algorithm,” UAI, 2007
9. Hartley, Rao, “Classification and estimation in analysis of variance problems”, Rev.Int.Stat.Inst., 1968
10. Healy, Westmacott, “Missing values in experiments analysed on automatic computers”, J.Roy.Stat.Soc., 1956
11. Joachims, “Transductive inference for text classification using support vector machines”, ICML, 1999
12. Krijthe, Loog, “Optimistic semi-supervised least squares classification”, ICPR, 2016
13. Lafferty, Wasserman, “Statistical analysis of semi-supervised regression”, NIPS, 2007
14. Loog, Duin, “The dipping phenomenon”, S+SSPR 2012
15. Loog, Jensen, “Semi-Supervised Nearest Mean Classification Through a Constrained Log-Likelihood”, IEEE TNNLS, 2015
16. McLachlan, “Iterative reclassification procedure for constructing an asymptotically optimal rule (...)”, JASA, 1975
17. McLachlan, Krishnan, “The EM algorithm and extensions”, Wiley, 2008
18. Nagy, Shelton, “Self-corrective character recognition system,” IEEE TIT, 1966
19. Seeger, “Learning with labeled and unlabeled data”, technical report, 2001
20. Singh, Nowak, Zhu, “Unlabeled data: now it helps, now it doesn't”, NIPS, 2008
21. Sokolovska, Cappé, Yvon, “The asymptotics of semi-supervised learning”, ICML, 2008
22. Yates, “The analysis of replicated experiments when the field results are incomplete”, Emp.J.Exp.Agric., 1933
23. Zhou, Bousquet, Lal, Weston, Schölkopf, “Learning with local and global consistency”, NIPS, 2004
24. Zhu, “Semi-supervised learning literature survey”, University of Wisconsin, TR 1530, 2008
25. Zhu, Ghahramani, Lafferty, “Semi-supervised learning using Gaussian fields and harmonic functions”, ICML, 2003