# Safety / Pessimism / Contrast

Marco Loog /
Jesse H. Krijthe

KU /
TU
Delft

- Qs about earlier lecture?  Mine or others'

- Recap : why do EM and related approaches fail?
  - Lots of unlabeled data -> always [or at least often] converge to same solution
  - "Problem" is that people label objects
  - Always same solution cannot always be right...

- More generally : wrong assumptions may get you in trouble

- Always have to make some assumptions on the data
  - Or do we?
  - Realize that choice of classifier is already an assumption
  - So can we leverage that?
  - Let us study if and in what sense this might be possible...

- Firstly, we look at performance on the data at hand

- Secondly, we only consider strictly safe techniques
  - Those that never get worse than the supervised version for every instantiation of the data!
  - Fairly strong assumption [why would we want this?]
  - And worse in what sense? $L(\theta_{\text{semi}}) < L(\theta_{\text{sup}})$

- Which classifier is trained such that it optimizes the error rate?
  - They don’t exist really / they typically are not considered

- Also, dipping basically shows there is little hope that, generally, we can improve in terms of error rate
  - Or accuracy or AUC etc.
  - One can indeed proof that such is impossible
  - But if a classifier optimizes particular loss, we can at least study how that loss behaves in semi-supervised setting

- I sketch and discuss two particular constructs to obtain insight in this matter

- Projection estimators
  - Sketch the general idea
  - Also geometrically
  - Try to get the gist across

- Minimax estimators
  - Basically equivalent to projection estimation
  - Start from true risk and introduce empirical risk
  - Consider the two-class setting
  - Make clear what is the difference between supervised and semi-supervised settings
  - How to estimate? : contrast and pessimism

- Make clear that difference will never be larger than 0
  - Discuss when do we have strict inequality?
  - And what does it mean if the inequality is not strict
  - What does it mean if strictness is not possible?

- So, what can we show in the end?

  - Safety is possible for classical classifiers like NMC, LDA and QDA, for classifiers relying on L1 and L2 losses, and for many generative models

  - It is not possible for SVMs and logistic regression [how very unfortunate]

- In conclusion?
  - In some cases, we can do a lot; give strong guarantees
  - But in others, it seems there is little SSL can deliver
  - Realize, however, that we considered the rather stringent setting of safe SSL for every data instantiation

- Three things to consider
  - It might work for non-loss based classifiers
  - More general theoretical headway is still possible : e.g. we could ask for “safety in the average”
  - If we can make fair additional assumptions on the data — if we understand the data we are dealing with — unlabeled data can help a lot [but don’t blindly trust the tools you use and study your data!]

- Qs?

## ▪ A nicely biased list of references


1. Ben-David, Lu, Pal, “Does unlabeled data provably help?”, COLT, 2008
2. Cohen, Cozman, Sebe, Cirelo, Huang, “Semisupervised learning of classifiers,” IEEE TPAMI, 2004
3. Cozman, Cohen, “Risks of semi-supervised learning,” in Semi-Supervised Learning. MIT Press, 2006
4. Kawakita, Takeuchi, “Safe semi-supervised learning based on weighted likelihood,” NN, 2014
5. Krijthe, Loog, “Robust Semi-supervised Least Squares Classification by Implicit Constraints”, arXiv, 2015
6. Krijthe, Loog, “Projected Estimators for Robust Semi-supervised Classification”, arXiv, 2016
7. Krijthe, Loog, “Optimistic semi-supervised least squares classification”, ICPR, 2016
8. Krijthe, Loog, “The pessimistic limits of margin-based losses in semi-supervised learning”, as yet unpublished, 2016
9. Lafferty, Wasserman, “Statistical analysis of semi-supervised regression”, NIPS, 2007
10. Li, Zhou, “Towards making unlabeled data never hurt”, IEEE TPAMI, 2015
11. Loog, “Constrained parameter estimation for semi-supervised learning: the case of the nearest mean classifier”, ECML, 2010
12. Loog, “Semi-supervised linear discriminant analysis through moment-constraint parameter estimation”, PRL, 2014
13. Loog, “Contrastive Pessimistic Likelihood Estimation for Semi-Supervised Classification”, IEEE TPAMI, 2016
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. Loog, Krijthe, Jensen, “On measuring and quantifying performance”, Ch1.3 in Handbook of PR and CV, 2016
17. Poggio, Smale, “The mathematics of learning”, Notices of the AMS, 2003
18. Seeger, “Learning with labeled and unlabeled data”, technical report, 2001
19. Singh, Nowak, Zhu, “Unlabeled data: now it helps, now it doesn't”, NIPS, 2008
20. Sokolovska, Cappé, Yvon, “The asymptotics of semi-supervised learning”, ICML, 2008