Normal Deviate

Thoughts on Statistics and Machine Learning

Guest Post: ROB TIBSHIRANI

GUEST POST: ROB TIBSHIRANI

Today we have a guest post by my good friend Rob Tibshirani (http://www-stat.stanford.edu/~tibs/). Rob has a list of nine great statistics papers. (He is too modest to include his own papers.) Have a look and let us know what papers you would add to the list. And what machine learning papers would you add? Enjoy.

9 Great Statistics papers published after 1970
Rob Tibshirani

I was thinking about influential and awe-inspiring papers in Statistics and thought it would be fun to make a list. This list will show my bias in favor of practical work, and by its omissions, my ignorance of many important subfields of Statistics. I hope that others will express their own opinions.

1. *Regression models and life tables (with discussion)* (Cox 1972). A beautiful and elegant solution to an extremely important practical problem. Has had an enormous impact in medical science. David Cox deserves the Nobel Prize in Medicine for this work.
2. *Generalized linear models* (Nelder and Wedderburn 1972). Formulated the class of generalized regression models for exponential family distributions. Provided the framework for the {\tt glim} package and the S and R modelling languages.
3. *Maximum Likelihood from Incomplete Data via the {EM} Algorithm (with discussion)* (Dempster, Laird, and Rubin 1977). Brought together many related ideas for dealing with missing or messy data, in one conceptually simple and powerful framework.
4. *Bootstrap methods: another look at the jackknife* (Efron 1979). Introduced one of the first computer-intensive statistical tools. Widely used in many scientific fields
5. *Classification and regression trees* (Breiman, Friedman, Olshen and Stone 1984). Not a paper, but a book. Among the first proposals for data mining to demonstrate the power of a detailed practical implementation of a method, including cross-validation for model selection
6. *How biased is the error rate of a prediction rule?* (Efron 1986). Greatly advanced our understanding of training and test error rates, and overfitting and ways to deal with them.

7. *Sampling based approaches to calculating marginal densities* (Gelfand and Smith 1990). Buidling on earlier work by Geman and Geman, Tanner and Wong, and others, this paper developed a simple and elegant sampling-based method for estimating marginal densities. Huge impact on Bayesian work

8. *Controlling the false discovery rate: a practical and powerful approach to multiple testing* (Benjamini and Hochberg 1995). Introduced the FDR and a selection procedure whose FDR is controlled at a given level. Enormously influential in the modern age of high-dimensional data.

9. *A decision-theoretic generalization of online learning and an application to boosting* (Freund and Schapire 1995). Not a statistics paper per se, but one that introduced one of the most powerful supervised learning methods and changed the way that many of us thought about the prediction problem.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B.*, 85, 289-300.

Breiman, L. and Friedman, J. and Olshen, R. and Stone, C. (1984). *Classification and Regression Trees*, Wadsworth, New York.

Cox, D.R. (1972). Regression models and life tables (with discussion). *J. Royal. Statist. Soc. B.*, 74, 187-220.

Dempster, A., Laird, N and Rubin, D. (1977). Maximum Likelihood from Incomplete Data via the {EM} Algorithm (with discussion). *Journal of the Royal Statistical Society Series B*, 39, 1-38.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, 7, 1-26.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81, 461-470.

Freund, Y. and Schapire, R. (1997). A decision-theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences*, 55, 119-139.

Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Nelder, J.A. and Wedderburn, R.W. (1972). Generalized linear models. *J. Royal Statist. Soc. B.*, 135, 370-384.

This entry was written by <u>normaldeviate</u>, posted on December 21, 2012 at 9:57 am (2012-12-21T09:57:52-0400), filed under <u>Uncategorized</u>. Bookmark the <u>permalink</u>. Follow any comments here with the <u>RSS feed for this post</u>. Both comments and trackbacks are currently closed.

11 Comments

1. Antoine
   Posted December 21, 2012 at 12:56 pm | <u>Permalink</u>
   Another great paper from this period:

   – Akaike, Hirotugu (1974), "A new look at the statistical model identification", IEEE Transactions on Automatic Control 19 (6): 716–723

2. oz
   Posted December 21, 2012 at 3:06 pm | <u>Permalink</u>
   Agree with the above on AIC. Also the BIC paper: (possibly the most influential perword/page)
   G Schwarz (1978) – Estimating the dimension of a model, The annals of statistics, 6(2), p. 461-464.

   Some other personal favorites:
   1. The paper introducing support-vector-machines:
   C Cortes, V Vapnik (1995) – Support-vector networks, Machine learning, 20(3) p. 273-297
   2. The lasso paper (omitted from above due to Rob's modesty):
   R Tibshirani (1996) – Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological), 267-288
   3. The LDA paper (since any 'greatest' list wouldn't be complete without Michael Jordan):
   D Blei, A Ng and M Jordan (2003) – Latent Dirichlet allocation. Journal of Machine Learning Research 3 (4–5): p. 993–1022

3. <u>adrianpradana</u>
   Posted December 21, 2012 at 7:03 pm | <u>Permalink</u>

Reblogged this on <u>adrianpradana</u>.

4. CC
   Posted December 22, 2012 at 12:46 am | <u>Permalink</u>
   Hey I was trying to find the "Generalized linear models" paper by Nelder and Wedderburn
   by searching in Journal of Royal Statistical Society. Series B (Methodological) but couldn't
   find it, that was because it's actually in Series A (General) instead. Just a heads up in case
   you want to fix it.

5. Jonathan Goodman
   Posted December 22, 2012 at 9:21 am | <u>Permalink</u>
   I don't agree with listing the work of Geman Geman, Gelfand and Smith. All they did was
   learn existing Monte Carlo methods. And they didn't do very well at that. What they call
   "Gibbs sampler" had been used by computational physicists and chemists for decades
   under names like "heat bath" and "partial resampling".

6. Umberto Picchini
   Posted December 22, 2012 at 5:05 pm | <u>Permalink</u>
   I would like to add:
   1) J. Friedman (1991). Multivariate adaptive regression splines (with Discussion), The Annals
   of Statistics 19(1), pp 1-141.
   2) the book by J. Pinheiro and D. Bates (2000), Mixed-effects models in S and S-Plus,
   Springer. It has been a very important reference for S and R users and pushed practical
   application of hierarchical modelling beyond NONMEM.
   3) together with my suggestion (2) above we should add some article on mixed-effects
   modelling, but I am not sure which one would be best. Perhaps Laird, Nan M.; Ware, James
   H. (1982). "Random-Effects Models for Longitudinal Data", Biometrics 38(4), 963-974.

7. Jason
   Posted December 23, 2012 at 2:26 pm | <u>Permalink</u>
   In the statistics front:
   1) Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of
   the Royal Statistical Society. Series B (Methodological), 192-236.
   This paper contains many of the major ideas in the graphical models literature.
   2) I second the lasso paper by Rob. Also around the same time, the basis pursuit paper
   appeared in the Applied Math/Signal Processing literature: Chen, S. S., Donoho, D. L., &
   Saunders, M. A. (1998). Atomic decomposition by basis pursuit. SIAM journal on scientific
   computing, 20(1), 33-61.

8. <u>Tapen</u>
   Posted December 24, 2012 at 12:58 am | <u>Permalink</u>
   No time series? I would have thought Granger and Engle would feature somewhere.

9. <u>xi'an</u>
   Posted January 9, 2013 at 7:56 pm | <u>Permalink</u>

I started a PhD course on <u>reading classics</u> a few years ago, so that my students could learn to analyze and present papers, while getting a better vision of the broadness of the field. I will include Rob's suggestions that were not already in the intersection!

10. Radu
Posted January 17, 2013 at 6:37 pm | <u>Permalink</u>
I would add the bootstrap paper by Efron '79 and the paper by Hastings '70 where he introduces what we now call the Metropolis-Hastings sampler.

11. <u>epanechnikov</u>
Posted February 24, 2013 at 7:00 pm | <u>Permalink</u>
Reblogged this on <u>Epanechnikov's Blog</u> and commented:
9 Great Statistics papers published after 1970

3 Trackbacks

1. By <u>Linkage (technical) « An Ergodic Walk</u> on December 24, 2012 at 9:53 pm
[…] at Larry Wasserman's blog, Rob Tibshirani suggests 9 Great Statistics papers published after 1970. You know, in case you were looking for some light reading over winter […]

2. By <u>datanalytics » Los nueve artículos favoritos de Tibshirani</u> on January 8, 2013 at 2:20 am
[…] Tibshirani ha seleccionado recientemente una lista de nueve artículos de estadística publicados desde 1970. Son […]

3. By <u>Guest Post: ROB TIBSHIRANI « wendyranstat</u> on January 13, 2013 at 9:47 am
[…] Guest Post: ROB TIBSHIRANI. […]

<u>Blog at WordPress.com.</u> | <u>The Sandbox 1.6.2 Theme</u>.