Stat 687 - A sample report, and course requirements

Rebecka Jörnsten*
Department of Statistics
Rutgers University
Piscataway, NJ 08854, USA

Abstract

The lectures for this class will consist of informal discussions of papers on various topics. Each student will be assigned a paper at the onset of the semester. The first task is to write a brief summary of this paper (3-5 pages, more detailed instructions below). In addition, you will prepare a 30 minute presentation for the class. Take the opportunity to practise speaking about a topic, and organizing a talk and a report. You will get useful feedback from the faculty in the audience, as well as your fellow students.
All students must submit a set of questions (3 minimum) and comments on the each preliminary student presentation. These questions can concern the topic of the paper: "Please explain how method X relates to method Y - I thought this was unclear in your presentation", "You mentioned that method X is used for Z, but couldn't it also be useful for W?". Comments on the presentation itself should be constructive: "I think that the second table was a bit messy - can it be turned into a figure? or split in two?", "When you talked about method X you looked down the whole time, and it made it difficult to hear what you were saying. Perhaps you can pick a person/friend in the audience to look at while you're talking?"
The second task is to take the comments you received from the preliminary presentation, and follow-up on the paper you presented. What advancements did this paper lead to? Are there open questions still? In what kind of applications do you see these methods used? Do a literature review, and address the questions/comments from the class. Prepare a final report ($\sim$ 10 pages) and a final presentation (40 minutes). You will hand in the final report at the end of the semester.

1 Introduction

This class has many goals: (1) Practise reading about an unfamiliar topic, and critically thinking about it, (2) Literature review - how to learn more about a topic, (3) Verbal presentation skills - practise summarizing information and communicating it to an audience, (4) Writing presentation skills - practise summarizing information in a report format.

As mentioned in the (somewhat lengthy) abstract, there are several requirements for this class. They may seem daunting at first, but remember - we are offering this class to help you improve your presentation skills, and to give you feedback. It is important to be able to communicate your results to an audience, both verbally and in writing.

*contact: rebecka@stat.rutgers.edu, fax +1 732 445-3428, telephone +1 732 445-3145

Required Tasks

1	Short presentation	30 minutes
2	Paper summary	3-5 pages Synopsis, criticism
3	Comments	3 minimum
4	Final presentation	40 minutes Incorporate comments Literature review
5	Final report	10 pages

2 Methods

This is a latex template for a report. I will post it at
`http://www.stat.rutgers.edu/řebecka/Stat687`
if you want to use it.

Here are some suggestions for your reports.

2.1 Structuring your report

It is difficult for a reader to follow if you don't bin things up a bit. Use separate sections headings, and think carefully about the order in which you introduce the material.

For the first report, the paper summary, use the abstract to clearly outline the main contribution of the paper. You may also include a particular question you have in the abstract.

The introduction can be used to briefly summarize the paper. Give some background - why is this an important topic? For the second paper, you will need to beef up the introduction. You want to give a time-line for the papers you reviewed, or for how the method was developed/improved.

Since these reports are not a research paper, the structure will be somewhat different. However, you can mimic the structure of a research paper by adding a methods section after the introduction. In this section, outline the assumptions and the statistical methodologies of the paper(s) you reviewed. Make sure to not simply copy the paper you read. Rephrase in your own words as much as possible, and add explanations where you feel they are needed. Perhaps you can come up with a slightly different explanation/illustration from the authors?

After the methods section you add a discussion. In the discussion you can again highlight the main contribution of the paper(s) you reviewed. You can now also put this is a larger context: where do you think this method is most useful? You can add your own comments here as well: was there something about the paper you particularly liked/something that was difficult to follow/something that surprised you. This discussion should not be too superficial (e.g. I would have used the word "each" instead of "every", and I would have used color figures...). Address the material at hand.

2.2 Appearance

It would be great if results could stand on their own merit, but they can't. You have to help the reader appreciate the results by wrapping them up nicely. Here are a few suggestions:

- Use full sentences, proper punctuation, and run spell check

- Use sections, sub-sections, and refer to the sub-sections from other parts of the document
- Be consistent in terms of notation and writing style
- Figures and tables should be clearly labeled and the captions should be informative (not just a table/figure number).
- Don't contradict yourself. Don't copy the authors or any other text.
- Provide a complete list of references (see sample below).

3 Discussion

Finally, we hope you will enjoy this class. The format is an informal discussion class, but do make an effort with your presentations and reports. This way, the feedback we give you will more likely reflect the help you truly need.

4 Suggested reading list

(I) Subject: Bagging
1) Leo Breiman, Bagging predictors
http://www.springerlink.com/content/r532018156xp344g/fulltext.pdf
2) Peter Buhlmann, Bin Yu, Analyzing Bagging
http://projecteuclid.org/Dienst/UI/1.0/Summarize/euclid.aos/1031689014

(II) Subject: Cox regession model
1) Cox, D.R. (1972). Regression models and life tables. *J. Royal Statist. Soc. B* **34** 187-220.

(III) Subject: EM-algorithm
1) Dempster, A.P., Larid N.M. and Rubin, D.B. (1977). Maximum Likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. B* **39** 1-38.
2) Meng, X. L., van Dyk, D. A. (1997). "The EM algorithm - an old folk song sung to a fast new tune." Read paper in J. R. Statist. Soc., Ser. B, 59, 511-567.

(IV) Subject: Extreme value theory
1) L. de Haan and J. de Ronde (1998). Sea and wind: multivariate extremes at work. **Extremes**, 1, 7-45.

(V) Subject: Empirical Bayes
1) Robbins, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157-163. Univ. of California Press, Berkeley.
2) Stein, C. (1981) "Estimation of the mean of a multivariate normal distribution", Ann Statist., 9, 1135-1151.
3) Berger, J. (1975), "Minimax estimation of location vectors for a wide class of densities", Ann. Statist. 4, 33-50.
4) Casella, G¿ and Strawderman, W.E. (1981), "Estimating a bounded normal mean", Ann. Statist., 9, 870-878.
5) Hwang, J. T., (1982), "Improving on standard estimators in discrete exponential families with applications to Poisson and negative binomial distributions", Ann. Statist.,10, 857-867.

(VI) Subject: Dimension reduction
1) P.J. Huber, Projection Pursuit, The Annals of Statistics, vol. 13, no. 2, pp. 435-475, 1985. Discussion paper

2a) "Independent Component Analysis: A Tutorial" Aapo Hyvrinen and Erkki Oja Helsinki University of Technology Laboratory of Computer and Information Science http://www.cs.helsinki.fi/u/ahyvarin/papers/NN00new.pdf
2b) "Survey on Independent Component Analysis" Aapo Hyvrinen http://www.cis.hut.fi/ aapo/ps/NCS99.pdf Helsinki University of Technology Laboratory of Computer and Information Science

(VII) Subject: Smoothing
1) Chu, C.K., Glad, I.K., Godtliebsen, F., Marron, J.S., 1998. "Edge-preserving smoothers for image processing." Journal of the American Statistical Association 93, 526-541. Discussion paper.

(VIII) Subject: Power-1 test
1) Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41** 1397-1409.

References

[Banfield and Raftery, 1993] Banfield, J. D., Raftery, A. E. (1993) Model-based Gaussian and non-Gaussian clustering, *Biometrics* **49**: 803-821.

[Bar-Joseph *et al.*, 2003] Bar-Joseph,Z., Gerber,G., Gifford,D., Jaakkola,T., Simon,I. (2003) Continuous representations of time-series gene expression data, *J Comput Biol* **10**(3): 341-56.

[Bø *et al.*, 2004] Bø,T., Dysvik,B., Jonassen,I. (2004) LSimpute: accurate estimation of missing values in microarray data with least squares methods, *Nucleic Acids Res* **32**(3): e34.

[Brown and Botstein, 1999] Brown,P., Botstein,D. (1999) Exploring the new world of the genome with DNA microarrays, *Nat Genet*, **21**, 33-7.

[Chen *et al.*, 2002] Chen,X., Cheung,S., So,S., Fan,S., Barry,C., Higgins,J., Lai,K., Ji,J., Dudoit,S., Ng,I., Van,D., Rijn,M, Botstein,D., Brown,P. (2002) Gene expression patterns in human liver cancers, *Mol Biol Cell* **13**(6): 1929-39.

[Churchill, 2002] Churchill,G.A. (2002) Fundamentals of experimental design for cDNA microarrays, *Nature Genetics* **32** Suppl: 490-5.

[Cui and Churchill, 2003] Cui,X., Churchill,G.A. (2003) Statistical tests for differential expression in cDNA microarray experiments, *Genome Biology* **4**: 210.

[Dempster *et al.*, 1977] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. R. Stat. Soc. B* **39**(1): 1-38.

[Eisen *et al.*, 1998] Eisen,M., Spellman,P., Brown,P., Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns, *Proc Natl Acad Sci U S A*, **95**, 14863-8.

[Jörnsten *et al.*, 2005] Jörnsten,R., Wang,H.Y., Welsh,W.J., Ouyang,M. (2005) DNA microarray data imputation and significance analysis of differential expression, *Bioinformatics* **21**(22): 4155-61.

[Kim *et al.*, 2005] Kim,H., Golub,G.H., Park,H. (2005) Missing value estimation for DNA microarray gene expression data: local least squares imputation, *Bioinformatics* **21**(2): 187-98.

[Kim *et al.*, 2004] Kim,K.Y., Kim,B.J., Yi,G.S. (2004) Reuse of imputed data in microarray analysis increases imputation efficiency, *BMC Bioinformatics* **5**:160.