MATH 6267 SYLLABUS

Spring 2013

Course Number:	Math 6267
Course Name:	Multivariate Statistical Analysis Standard setting and High-dimensional theory
Lecture Time:	MW 3:05-4:25 pm
Lecture Room:	169
Instructor:	Pr. Karim Lounici Office: Skiles 228 Email Address: `klounici AT math.gatech.edu`
Course Web Page:	`http://www.math.gatech.edu/~klounici6/6267/6267.html`
Office Hours:	MW 2:00-3:00pm and by appointment.
Contacting me:	You can contact me by email.
Prerequisites:	A solid background in Mathematics and Statistics/Probability is required. Math 4261, Math 4262 and Math 6241

Course Objective: The first part of the course concerns the classical multivariate theory. The goal is to present standard results and techniques on multivariate gaussian vectors, linear regression and reduction of dimension. The second part of the course treats very recent developments on high dimensional statistics. The goals are to become familiar with the most important and up to date methods of variables selection and aggregation in the field of high-dimensional statistics, to understand their theorethical and computational merits and limitations in a large array of statistical models, e.g., parametric and non parametric regression, classification, density estimation.

Textbook: The first part of the course will be based on the two following references:

1. Mardia, Kent and Bibly *Multivariate Analysis*
2. Anderson *An Introduction to Multivariate Statistical Analysis*

The second part of the course is based on very recent research and there is not yet any textbook that covers it whole. A comprehensive list of research papers and related books is provided at the end of this syllabus.

Course Outline:

Below is a tentative outline of the topics we will treat in this class.

1. **Multivariate Gaussian vectors:** Multivariate normal distribution theory, chi-square, Wishart distribution, Cochran's theorem.
2. **Principal Component Analysis:** Dimension reduction principle, applications, asymptotic distribution.
3. **Discriminant analysis and Classification:** Discrimination and classification principle, linear classification, Bayes rule, error of classification.
4. **Introduction to high-dimensional statistics:** Motivation and challenges. Prediction, estimation and variables selection Problems. high-dimensionality and computational tractability.
5. **Lasso, Dantzig selector, Group Lasso, Exponential Weights:** Sparsity oracle inequalities. Non asymptotic variables selection. Standard and structure sparsity. Algorithms. Tuning of the regularization parameter. Minimax and non minimax lower bounds.

Grading Policies: There will be a maximum of three homeworks, tests or report on research papers with oral presentation in class. Tests are to be taken without any notes of any sort. Late homeworks will NOT be accepted for grading (justified emergencies excepted). Discussing the problems with everyone is encouraged but everyone needs to hand in a personal copy. Homework may require using R software (this is a free software and a short tutorial will be provided).
Letter grades will be based on the accumulated points according to the standard cutoffs: A:90-100, B:80-89, C:70-79, D:60-69, F:0-59.

Honor Code: All students are expected to comply with the Georgia Tech Honor Code. Any violations of the Georgia Tech Honor Code will be submitted directly to the Dean of Students. The Georgia Tech Honor Code is available at http://www.honor.gatech.edu/

References

[1] F. Bach. Consistency of the group Lasso and multiple kernel learning. *Journal of Machine Learning Research*, 9:1179–1225, 2008.

[2] P.J. Bickel, Y. Ritov, and A. Tsybakov. Simultaneous analysis of lasso and dantzig selector. *Annals of Statistics*, 37(4):1705–1732, 2009.

[3] O. Bousquet. A Bennett concentration inequality and its application to suprema of empirical processes. *C. R. Math. Acad. Sci. Paris*, 334(6):495–500, 2002. ISSN 1631-073X.

[4] F. Bunea. Consistent selection via the Lasso for high dimensional approximating regression models. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, volume 3 of *Inst. Math. Stat. Collect.*, pages 122–137. Inst. Math. Statist., Beachwood, OH, 2008.

[5] F. Bunea, A.B. Tsybakov, and M.H. Wegkamp. Sparsity oracle inequalities for the Lasso. *Electronic Journal of Statistics*, 1:169–194, 2007.

[6] F. Bunea, A.B. Tsybakov, and M.H. Wegkamp. Aggregation for Gaussian regression. *Annals of Statistics*, 35:1674–1697, 2007.

[7] E. Candes and T. Tao. The Dantzig selector: Statistical estimation when p is much larger than n. *Annals of Statistics*, 35(6):2313–2351, 2005.

[8] O. Catoni. *Statistical learning theory and stochastic optimization*, volume 1851 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2004. ISBN 3-540-22572-2. Lecture notes from the 31st Summer School on Probability Theory held in Saint-Flour, July 8–25, 2001.

[9] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, Cambridge, 2006. ISBN 978-0-521-84108-5; 0-521-84108-9.

[10] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159 (electronic), 2001. ISSN 0036-1445. Reprinted from SIAM J. Sci. Comput. **2**0 (1998), no. 1, 33–61 (electronic) [MR1639094 (99h:94013)].

[11] A.S. Dalalyan and A.B. Tsybakov. Aggregation by exponential weighting and sharp oracle inequalities. In *Learning theory*, volume 4539 of *Lecture Notes in Comput. Sci.*, pages 97–111. Springer, Berlin, 2007.

[12] A.S. Dalalyan and A.B. Tsybakov. Pac-bayesian bounds for the expected error of aggregation by exponential weights. Technical report, Universitï¿$\frac{1}{2}$ Paris 6, CREST and CERTIS, Ecole des Ponts ParisTech, 2009. personal communication.

[13] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Ann. Statist.*, 32(2):407–499, 2004. ISSN 0090-5364. With discussion, and a rejoinder by the authors.

[14] I.E. Franck and J.H. Friedman. A statistical point of view of some chemometrics regression tools (with discussion). *Technometrics*, 35:109–148, 1993.

[15] J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. Technical report, Irisa, Universitï¿$\frac{1}{2}$ de Rennes, 2004.

[16] A. Juditsky and A. Nemirovski. On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization. *submitted to Mathematical Programming Series B., Special Issue on Machine Learning*, 2008. arXiv:0809.2650v1.

[17] A. Juditsky, P. Rigollet, and A. B. Tsybakov. Learning by mirror averaging. *Ann. Statist.*, 36(5):2183–2206, 2008. ISSN 0090-5364.

[18] K. Knight and W. Fu. Asymptotics for lasso-type estimators. *Ann. Statist.*, 28(5): 1356–1378, 2000. ISSN 0090-5364.

[19] V. Koltchinskii. Sparsity in penalized risk minimization. *Ann. IHP*, to appear.

[20] V. Koltchinskii. Dantzig selector and sparsity oracle inequalities. *Bernoulli*, to appear.

[21] V. Koltchinskii and M. Yuan. Sparse recovery in large ensembles of kernel machines. In *Conference on Learning Theory, COLT-2008*, pages 229–238, 2008.

[22] G. Lecué. Lower bounds and aggregation in density estimation. *J. Mach. Learn. Res.*, 7:971–981, 2006. ISSN 1532-4435.

[23] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. ISBN 3-540-52013-9. Isoperimetry and processes.

[24] G. Leung and A.R. Barron. Information theory and mixing least-squares regressions. *IEEE Trans. Inform. Theory*, 52(8):3396–3410, 2006. ISSN 1557-9654.

[25] K. Lounici. Generalized mirror averaging and D-convex aggregation. *Math. Methods Statist.*, 16(3):246–259, 2007. ISSN 1066-5307.

[26] K. Lounici. Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electronic Journal of Statistics*, 2:90–102, 2008.

[27] K. Lounici, M. Pontil, A.B. Tsybakov, and S.A. van de Geer. Taking advantage of sparsity in multi-task learning. *COLT (electronic)*, 2009.

[28] N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. *Ann. Statist.*, 34(3):1436–1462, 2006. ISSN 0090-5364.

[29] A. Nemirovski. Topics in non-parametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, volume 1738 of *Lecture Notes in Math.*, pages 85–277. Springer, Berlin, 2000.

[30] G. Obozinski, M.J. Wainwright, and M.I. Jordan. Union support recovery in high-dimensional multivariate regression. Technical report, Berkeley Stat., 2008.

[31] Ph. Rigollet and A. B. Tsybakov. Linear and convex aggregation of density estimators. *Math. Methods Statist.*, 16(3):260–280, 2007. ISSN 1066-5307.

[32] M. Rosenbaum and A.B. Tsybakov. Sparse recovery under matrix uncertainty. *Ann. Statist., submitted*, 2009.

[33] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996. ISSN 0035-9246.

[34] A.B. Tsybakov. Optimal rates of aggregation. In *Computationnal Learning theory and Kernel Machines (COLT)*, volume 2777 of *Lecture Notes in Artificial Intelligence*, pages 303–313. Springer, heidelberg, 2003.

[35] A.B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *Ann. Statist.*, 32(1):135–166, 2004. ISSN 0090-5364.

[36] Alexandre B. Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004. ISBN 3-540-40592-5.

[37] S.A. van de Geer. The deterministic lasso. Technical report, Seminar für Statistik ETH, Zürich, 2007.

[38] A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3.

[39] M. Waingwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using l_1-constrained quadratic programming (lasso). *IEEE*, 2009.

[40] M. Wegkamp. Lasso type classifiers with a reject option. *Electron. J. Stat.*, 1:155–168 (electronic), 2007. ISSN 1935-7524.

[41] Y. Yang. Combining different procedures for adaptive regression. *J. Multivariate Anal.*, 74(1):135–161, 2000. ISSN 0047-259X.

[42] Y. Yang. Aggregating procedures for a better performance. *Bernoulli*, 10:25–47, 2004.

[43] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 68 (1):49–67, 2006.

[44] H. Zou. The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.*, 101(476): 1418–1429, 2006. ISSN 0162-1459.