**List of papers for reading in spring class**

1. **Efron:** **statistical curvature, bootstrap**

Efron, Bradley. "Defining the curvature of a statistical problem (with applications to second order efficiency)." *The Annals of Statistics* (1975): 1189-1242.

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

1. **Missing data: EM paper, Wu’s proof (in one lecture I will do)**

Wu, CF Jeff. "On the convergence properties of the EM algorithm." *The Annals of Statistics* (1983): 95-103.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 39.1 (1977): 1-22.

1. **Causality**

Holland, Paul W. "Statistics and causal inference." *Journal of the American Statistical Association* 81.396 (1986): 945-960.

Rubin, Donald B. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66.5 (1974): 688-701.

Rubin, Donald B. "For objective causal inference, design trumps analysis." *The Annals of Applied Statistics* 2.3 (2008): 808-840.

Rosenbaum, Paul R., and Donald B. Rubin. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70.1 (1983): 41-55.

1. **Sliced inverse regression and extension**

Li, Ker-Chau. "Sliced inverse regression for dimension reduction." *Journal of the American Statistical Association*86.414 (1991): 316-327.

Cook, R. Dennis, and Sanford Weisberg. "Sliced inverse regression for dimension reduction: Comment." *Journal of the American Statistical Association* 86.414 (1991): 328-332.

Bura, Efstathia, and R. Dennis Cook. "Extending sliced inverse regression: The weighted chi-squared test." *Journal of the American Statistical Association* 96.455 (2001): 996-1003.

(one presentation should include all three above)

Cook, R. Dennis. "Principal Hessian directions revisited." *Journal of the American Statistical Association* 93.441 (1998): 84-94.

1. **biostatistics**

Cox, David R. "Regression models and life‐tables." *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (1972): 187-202.

Liang, Kung-Yee, and Scott L. Zeger. "Longitudinal data analysis using generalized linear models." *Biometrika* 73.1 (1986): 13-22.

Lai, Tze Leung, and Herbert Robbins. "Asymptotically efficient adaptive allocation rules." *Advances in Applied Mathematics* 6.1 (1985): 4-22.

1. **Functional data analysis**

Donoho, David L., and Iain M. Johnstone. "Adapting to unknown smoothness via wavelet shrinkage." *Journal of the American Statistical Association* 90.432 (1995): 1200-1224.

1. **Engr stat: Wu has 3-4 sets of slides.**

We have 3 videos by Prof. Jeff Wu.

Lai, Tze Leung. "Sequential changepoint detection in quality control and dynamical systems." *Journal of the Royal Statistical Society: Series B (Methodological)* 57.4 (1995): 613-644.

1. **Computer experiments: two major papers by O’Hagan (calibration plus Tuo-Wu; Multi-fidelity plus T-W-Yu)**

Kennedy, Marc C., and Anthony O'Hagan. "Bayesian calibration of computer models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.3 (2001): 425-464.

Kennedy, Marc C., and Anthony O'Hagan. "Predicting the output from a complex computer code when fast approximations are available." *Biometrika* 87.1 (2000): 1-13.

Tuo, Rui, and CF Jeff Wu. "Efficient calibration for imperfect computer models." *The Annals of Statistics* 43.6 (2015): 2331-2352.

Tuo, Rui, CF Jeff Wu, and Dan Yu. "Surrogate modeling of computer experiments with different mesh densities." *Technometrics* 56.3 (2014): 372-380.

**(Experimental design)**

* **(Latin hypercube sampling)** McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output. *Technometrics*, 21(2), 239-245.
* **(MaxPro)** Joseph, V. R., Gul, E., & Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, 102(2), 371-380.

1. **High-dimensional statistics**

Candès, Emmanuel J., Justin Romberg, and Terence Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information." *IEEE Transactions on Information Theo*ry 52.2 (2006): 489-509.

Candes, Emmanuel J., Justin K. Romberg, and Terence Tao. "Stable signal recovery from incomplete and inaccurate measurements." *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 59.8 (2006): 1207-1223.

(can choose only one of the two for class presentation)

1. **Machine learning (stat side)**

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996): 267-288.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (1995): 289-300.

(both very high impacts and citations)

**(Boosting)**

* **(AdaBoost)** Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.
* **(Gradient boosting)** Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.

**(MARS)** Friedman, J. H. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19(1), 1-67.

**(SVM)** Cortes, C. (1995). Support-Vector Networks. *Machine Learning*.

**(Approximation theorem of neural network)** Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators.*Neural networks*, 2(5), 359-366.

**(Spectral clustering)** Ng, A., Jordan, M., & Weiss, Y. (2001). On spectral clustering: analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14.

**(Independent component analysis)** Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks,* 13(4-5), 411-430.

1. **Donoho**

Donoho, David L. "Compressed sensing." *IEEE Transactions on Information Theory* 52.4 (2006): 1289-1306.

Donoho, David L., and Xiaoming Huo. "Uncertainty principles and ideal atomic decomposition." *IEEE Transactions on Information Theory* 47.7 (2001): 2845-2862. (Huo will present)

1. **MCMC**

Geman, Stuart, and Donald Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6 (1984): 721-741.

Green, Peter J. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." *Biometrika* 82.4 (1995): 711-732.

Gelfand, Alan E., and Adrian FM Smith. "Sampling-based approaches to calculating marginal densities." *Journal of the American Statistical Association* 85.410 (1990): 398-409.

1. **Bayes: computation outside MCMC**

Kass, Robert E., and Adrian E. Raftery. "Bayes factors." *Journal of the American Statistical Association* 90.430 (1995): 773-795.

1. **Stat optimization: EI paper**

Jones, Donald R., Matthias Schonlau, and William J. Welch. "Efficient global optimization of expensive black-box functions." *Journal of Global Optimization* 13.4 (1998): 455-492.

**15. Spatial Statistics**

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36(2), 192-225.

**16. Philosophy of Statistics**

(Two cultures in statistical modeling) Breiman, L. (2001). Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Statistical science*, 16(3), 199-231.

**17. Probability theory in queueing**

Dai, J. G. (1995). On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *Annals of Applied Probability*, 5(1), 49-77.

**18. Ridge regression**

Hoerl, Arthur E., and Robert W. Kennard. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics* 12.1 (1970): 55-67.

Students who choose Ridge Regression paper should also include below discussions:

Special issue on ridge regression, *Technometrics,* 2020.

Hoerl, Roger W. "Ridge regression: a historical context." *Technometrics* 62.4 (2020): 420-425.

(I will present the two above in early week)

Hastie, Trevor. "Ridge regularization: An essential concept in data science." *Technometrics* 62.4 (2020): 426-433. (interesting, can count as half presentation; or combined with lasso paper below)