15.097 Machine Learning via a Modern Optimization Lens

Place-Time: E51-345, TR: 2:30-4:00

Instructor:

Dimitris Bertsimas, E40-111; Tel.: (617) 253-4223; Office hours: by appointment; e-mail: dbertsim@mit.edu; homepage: http://web.mit.edu/dbertsim/www/

Teaching Assistants:

Colin Pawlowski, email: cpawlows@mit.edu, Office hours: TBA

Ying Daisy Zhuo, email: zhuo@mit.edu, Office hours: TBA

Recitation: Fridays: TBA.

Course Content and Objectives: The majority of the central problems of regression, classification and estimation have been addressed using heuristic methods even though they can be formulated as formal optimization problems. While continuous optimization approaches has had a significant impact in Machine Learning (ML)/ Statistics (S), mixed integer optimization (MIO) has played a very limited role, primarily based on the belief that MIO models are computationally intractable. The last three decades have witnessed a) algorithmic advances in MIO, which coupled with hardware improvements have resulted in an astonishing over 2 trillion factor speedup in solving MIO problems, b) significant advances in our ability to model and solve very high dimensional robust and convex optimization models.

Our objective in this course is to revisit some of the classical problems in ML/S and demonstrate that they can greatly benefit by a modern optimization treatment. The optimization lenses we use in this course are: convex, robust and mixed integer optimization. In all cases we demonstrate that optimal solutions to large scale instances (a) can be found in seconds, (b) can be certified to be optimal/near-optimal in minutes and (c) outperform classical heuristic approaches in out of sample experiments involving real and synthetic data.

The problems we address in this course include:

- variable selection in linear and logistic regression,
- convex, robust, and median regression,
- an algorithmic framework to construct linear and logistic regression models that satisfy properties of sparsity, robustness, significance, absence of multi-collinearity in an optimal way,
- optimal classification and regression trees and their relationship with neural networks,
- how to transform predictive algorithms to prescriptive algorithms,
- optimal prescriptive trees
- robust classification
- design of experiments via optimization
- missing data imputations using modern optimization,
- mixture of Gaussian models via MIO,
- exact bootstrap
- sparse matrix estimation including principal component analysis, factor analysis, inverse covariance matrix estimation and matrix completion.

Text: Research papers and class notes. All handouts can be downloaded from: https://stellar.mit.edu/S/course/15/sp15/15.097/

Recitations: The recitations will cover software implementation in Julia, computational aspects, and examples and applications that enhance the theory developed in the lectures.

Course Requirements: Problem sets, and one final team project. A project will need to involve up to two students per project. Grades will be determined by performance on the above requirements weighted approximately as 60% problem sets, and 40% final team project.

Lecture	Time	Topic	Readings
1	T, 2/06	The Optimization Lenses and Machine Learning	
2	R, 2/08	Best Subset in Linear Regression	[14, 27]
3	T, 2/13	Robust Linear Regression	[3]
4	R, 2/15	Algorithmic Framework for Linear Regression	[12, 19]
	T, 2/20	Class on Monday Schedule	
5	R, 2/22	Median Regression	[20]
6	T, 2/27	Convex Regression	[23]
7	R, 3/01	Classification: Sparsity and Robustness	[8, 13, 24]
8	T, 3/06	Support Vector Machines	[8]
9	R, 3/08	Optimal Classification Trees	[5]
10	T, 3/13	Optimal Regression Trees	[6]
11	R, 3/15	Optimal Prescriptive Trees	[7]
12	T, 3/20	Optimal Trees and Neural Networks	[21]
13	R, 3/22	Power of Optimization over Randomization	[9, 15]
	T, 3/27	Spring break	
	R, 3/29	Spring break	
14	T, 4/03	Identiying Exceptional Responders	[16]
15	R, 4/05	From Predictions to Prescriptions I	[10]
16	T, 4/10	From Predictions to Prescriptions II	[11, 22]
17	R, 4/12	Missing Data Imputations	[25]
	T, 4/17	Patriots Vacation–Vacation	
18	T, 4/19	Mixture of Gaussians via MIO	[1]
19	R 4/24	Exact Bootstrap	[26]
20	R, 4/26	Sparse Principal Component Analysis	[2]
21	T, 5/01	Certifiably Optimal Low Rank Factor Analysis	[4]
22	R 5/03	Certifiably Optimal Sparse Inverse Covariance Estimation	[17]
23	T, 5/08	Matrix Completion I	[28]
24	R, 5/10	Matrix Completion II	[18]
25	T, 5/15	Project Presentations	
26	R, 5/17	Project Presentations	

References

[1] H. Bandi, D. Bertsimas, and R. Mazumder. Learning a mixture of gaussians via mixed integer optimization. *INFORMS Journal of Optimization*, under review, 2018.

[2] L. Berk and D. Bertsimas. Certifiably optimal sparse principal component analysis. *Mathematical Programming Computation*, under review, 2017.

[3] D. Bertsimas and M. Copenhaver. Characterization of the equivalence of robustification and regularization in linear, median, and matrix regression. *European Journal of Operations Research, to appear*, 2017.

[4] D. Bertsimas, M. Copenhaver, and R. Mazumder. Certifiably optimal low rank factor analysis. *Journal of Machine Learning Research*, 18:1–53, 2017.

[5] D. Bertsimas and J. Dunn. Optimal trees. *Machine Learning*, 106(7):1039–1082, 2017.

[6] D. Bertsimas and J. Dunn. *Optimal Trees for Prediction and Prescription*. Dynamic Ideas, 2018.

[7] D. Bertsimas, J. Dunn, and N. Mundru. Optimal prescriptive trees. *INFORMS Journal of Optimization*, under review, 2018.

[8] D. Bertsimas, J. Dunn, C. Pawlowski, and Y. Zhuo. Robust classification. *INFORMS Journal of Optimization*, under review, 2017.

[9] D. Bertsimas, M. Johnson, and N. Kallus. The power of optimization over randomization in designing experiments involving small samples. *Operations Research*, 63 (4):868–876, 2015.

[10] D. Bertsimas and N. Kallus. From predictions to prescriptions. *Management Science*, under review, 2015.

[11] D. Bertsimas and N. Kallus. Pricing from observational data. *Management Science*, under review, 2017.

[12] D. Bertsimas and A. King. An algorithmic approach to linear regression. *Operations Research*, 64(1):2–16, 2016.

[13] D. Bertsimas and A. King. Logistic regression: From art to science. *Statistical Science*, 32(3):367384, 2017.

[14] D. Bertsimas, A. King, and R. Mazumder. Best subset selection via a modern optimization lens. *Annals of Statistics*, 44(2):813–852, 2016.

[15] D. Bertsimas, N. Korolko, and A. Weinstein. Covariate-adaptive optimization in online clinical trials. *Operations Research*, under review, 2017.

[16] D. Bertsimas, N. Korolko, and A. Weinstein. Identifying exceptional responders in randomized trials: An optimization approach. *INFORMS Journal on Optimization*, under review, 2018.

[17] D. Bertsimas and J. Lamperski. Certifiably optimal sparse inverse covariance estimation. *Mathematical Programming*, under reviewr, 2016.

[18] D. Bertsimas and M. Li. Sparse matrix completion. working paper, 2018.

[19] D. Bertsimas and M. Li. Systematic linear regression. working paper, 2018.

[20] D. Bertsimas and R. Mazumder. Least quantile regression via modern optimization. *Annals of Statistics*, 42 (6):2494–2525, 2014.

[21] D. Bertsimas, R. Mazumder, and M. Sobiesk. On the equivalence of neural networks and optimal trees. working paper, 2018.

[22] D. Bertsimas and C. McCord. From predictions to prescriptions in multistage optimization problems. *Mathematical Programming*, under review, 2017.

[23] D. Bertsimas and N. Mundru. Sparse convex regression. *INFORMS Journal on Computing*, under review, 2017.

[24] D. Bertsimas, J. Pauphilet, and B. van Parys. Sparse classification and phase transitions: a discrete optimization perspective. *Journal of Machine Learning Research*, under review, 2017.

[25] D. Bertsimas, C. Pawlowski, and Y. Zhuo. From predictive methods to missing data imputation: An optimization approach. *Journal of Machine Learning Research*, under review, 2017.

[26] D. Bertsimas and B. Sturt. Computation of exact bootstrap confidence intervals. *Operations Research*, under review, 2017.

[27] D. Bertsimas and B. van Parys. Sparse high dimensional regression: Exact scalable algorithms and phase transitions. *Annals of Statistics*, under review, 2016.

[28] R. Freund, P. Grigas, and R. Mazumder. An extended frank-wolfe method with "in-face" directions, and its application to low-rank matrix completion. working paper, 2015.