Computation & optimization for Lasso - part 2

Luyang Han & Janosch Ott

ETH Zurich

22 October 2018

Computation & optimization for Lasso - part 2

Luyang Han & Janosch Ott

ETH Zurich

22 October 2018

Overview

Overview

1. Coordinate Descent
2. A Simulation Study
3. Least Angle Regression
4. Digression: Duality
5. ADMM
6. Screening Rules

Digression: Duality in optimization

In various section, I came across terms like "dual" and "dual problem"

	Primal
Optimize	$\min f(x)$
Constraints	$g_i(x) \leq 0, h_j(x) = 0, x \in X$
Function	$L(x, \lambda, \mu) := f(x) + \sum_i \lambda_i g_i(x) + \sum_j \mu_j h_j(x)$
	Dual
Function	$q(\lambda, \mu) = \inf\limits_{x \in X} L(x, \lambda, \mu)$
Constraints	$\lambda \geq 0$
Optimize	$\max q(\lambda, \mu)$

Why though? - **Dual problem is always convex!**

Primal
Optimize $\min f(x)$
Constraints $g_i(x) \leq 0, h_j(x) = 0, x \in X$
Function $L(x, \lambda, \mu) := f(x) + \sum_i \lambda_i g_i(x) + \sum_j \mu_j h_j(x)$
Dual
Function $q(\lambda, \mu) = \inf\limits_{x \in X} L(x, \lambda, \mu)$
Constraints $\lambda \geq 0$
Optimize $\max q(\lambda, \mu)$
Why though? - **Dual problem is always convex!**

$x \in X$ for e.g. solutions in a cone or integer solutions
Terms: Primal problem, Lagrange function with dual variables/Lagrange-multipliers, dual function, dual problem
Dual problem is always convex! - I don't know much about optimization yet, but they really like convexity.
"(Convexity confers two advantages. The first is that, in a constrained problem, a convex feasible region makes it easier to ensure that you do not generate infeasible solutions while searching for an optimum.)
The second advantage is that all local optima are global optima. That allows local search algorithms to guarantee optimal solutions. And local search is often faster." [**?**])

References I

Trevor Hastie, Robert Tibshirani, and Martin Wainwright (2015)
Statistical learning with sparsity: the Lasso and generalizations
CRC Press; Boca Raton, FL

Jan De Leeuw (2015)
Block Relaxation Methods in Statistics
doi.org/10.13140/RG.2.1.3101.9607 (last accessed: 02.10.18)

S. Boyd
Alternating Direction Method of Multipliers
https://web.stanford.edu/~boyd/papers/pdf/admm_slides.pdf (last accessed: 14.10.18)

Geoff Gordon and Ryan Tibshirani (2012)
Uses of Duality
https://www.cs.cmu.edu/~ggordon/10725-F12/slides/18-dual-uses.pdf (last accessed: 14.10.18)

References I

Trevor Hastie, Robert Tibshirani, and Martin Wainwright (2015)
Statistical learning with sparsity: the Lasso and generalizations
CRC Press; Boca Raton, FL

Jan De Leeuw (2015)
Block Relaxation Methods in Statistics
doi.org/10.13140/RG.2.1.3101.9607 (last accessed: 02.10.18)

S. Boyd
Alternating Direction Method of Multipliers
https://web.stanford.edu/~boyd/papers/pdf/admm_slides.pdf (last accessed: 14.10.18)

Geoff Gordon and Ryan Tibshirani (2012)
Uses of Duality
https://www.cs.cmu.edu/~ggordon/10725-F12/slides/18-dual-uses.pdf (last accessed: 14.10.18)

References II

Paul Rubin (2016)
What are the advantages of convex optimization compared to more general optimization problems?
https://www.quora.com/
What-are-the-advantages-of-convex-optimization-compared-to-m
(last accessed: 14.10.18)

References II

Paul Rubin (2016)
What are the advantages of convex optimization compared to more general optimization problems?
https://www.quora.com/
What-are-the-advantages-of-convex-optimization-compared-to-m
(last accessed: 14.10.18)

Comments . . .

Questions . . .

Suggestions . . .

Comments . . .
Questions . . .
Suggestions . . .

That's it.
Thanks for listening.

Fill out your feedback sheets!

That's it.
Thanks for listening.

Fill out your feedback sheets!

You can test if **knitr** works with this minimal demo. OK, let's get started with some boring random numbers:

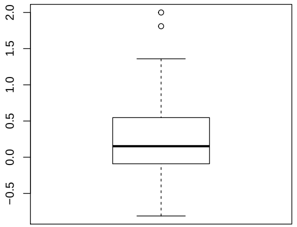
```
set.seed(1121)
(x=rnorm(20))

##  [1]  0.1449583  0.4383221  0.1531912  1.0849426  1.9995449 -(
##  [7]  0.1602680  0.5858923  0.3600880 -0.0253084  0.1508809  (
## [13]  1.3596812 -0.3269946 -0.7163819  1.8097690  0.5084011 -(
## [19]  0.1327188 -0.1559430

mean(x);var(x)

## [1] 0.3217385
## [1] 0.5714534
```

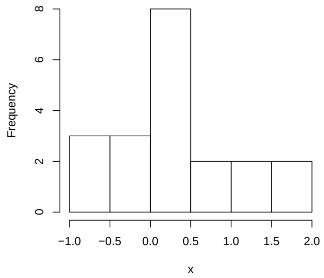

You can test if **knitr** works with this minimal demo. OK, let's get started with some boring random numbers:

```
set.seed(1121)
(x=rnorm(20))

##  [1]  0.1449583  0.4383221  0.1531912  1.0849426  1.9995449 -(
##  [7]  0.1602680  0.5858923  0.3600880 -0.0253084  0.1508809  (
## [13]  1.3596812 -0.3269946 -0.7163819  1.8097690  0.5084011 -(
## [19]  0.1327188 -0.1559430

mean(x);var(x)

## [1] 0.3217385
## [1] 0.5714534
```

The first element of x is 0.1449583. Boring boxplots and histograms recorded by the PDF device:

```r
## two plots side by side (option fig.show='hold')
boxplot(x)
hist(x,main='')
```





The first element of x is 0.1449583. Boring boxplots and histograms recorded by the PDF device:

two plots side by side (option fig.show='hold')
boxplot(x)
hist(x,main='')