Lecture 20 – Matrix optimization in ML

Primal-dual pair of problems

Primal problem

$$\max_{C \succ 0} \frac{m}{2}(\ln\det(C) - Tr(AC)) - \lambda \|C\|_1$$

Dual problem

$$\max_{W \succ 0}\{\frac{m}{2}\ln(\det(W)) - mp/2 : \text{ s.t. } \frac{m}{2}||(W - A)||_\infty \leq \lambda\}$$

Interior point method – $O(n^6)$ operations/iter

Block coordinate ascent

Update one row and one column of the dual matrix W at each step

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

$$\max_{W \succ 0} \{ \frac{m}{2} \ln(\det(W)) - mp/2 : \text{ s.t. } \frac{m}{2} ||W - A||_{\infty} \leq \lambda \}$$

$$\ln\det W = \ln(\det(W_{11})(w_{22} - {w_{12}}^T W_{11}^{-1} w_{12}))$$

Block coordinate ascent subproblem

Update one row and one column of the dual matrix W at each step

$$W = \begin{bmatrix} W_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}$$

$$\begin{array}{ll} \max_{w_{12}, w_{22}} & \ln(w_{22} - {w_{12}}^T W_{11}^{-1} w_{12})) \\ \text{s.t.} & \|w_{12} - a_{12}\|_\infty \leq \frac{2}{m}\lambda, |w_{22} - a_{22}| \leq \frac{2}{m}\lambda \end{array}$$

$$\min_{w_{12}} \{w_{12}^\top W_{11}^{-1} w_{12} : \quad \text{s.t.} \quad \|w_{12} - a_{12}\|_\infty \leq \frac{2}{m}\lambda,$$

Subproblem reformulation

$$\min_{w_{12}} \{ w_{12}^\top W_{11}^{-1} w_{12} : \quad \text{s.t.} \quad \|w_{12} - a_{12}\|_\infty \leq \frac{2}{m}\lambda,$$

$$w_{12} = W_{11}\beta$$

$$\min_{\beta} \{ \beta^\top W_{11} \beta : \quad \text{s.t.} \quad \|W_{11}\beta - a_{12}\|_\infty \leq \frac{2}{m}\lambda \}$$

Remember Lasso!

Primal-Dual pair of problems

$$\min \quad \frac{1}{2}||Ax - b||^2 + \lambda||x||_1$$

$$\min \quad \frac{1}{2}x^\top A^\top Ax$$
$$s.t. \quad \|A^\top(Ax - b)\|_\infty \leq \lambda$$

Dual subproblem

$$\min_{w_{12}}\{w_{12}^\top W_{11}^{-1} w_{12} : \quad \text{s.t.} \quad \|w_{12} - a_{12}\|_\infty \le \frac{2}{m}\lambda,$$

$$w_{12} = W_{11}\beta$$

$$\min_{\beta}\{\beta^\top W_{11}\beta : \quad \text{s.t.} \quad \|W_{11}\beta - a_{12}\|_\infty \le \frac{2}{m}\lambda\}$$

$$\min_{\beta}\{\|W_{11}^{1/2}\beta - W_{11}^{-1/2}a_{12}\|^2 + \frac{4}{m}\lambda\|\beta\|_1$$

The dual subproblem is the Lasso problem

Remember coordinate descent for Lasso

$$\min_{x_i} \quad \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1$$

Choose one variable x_i and column A_i.
Let $\bar{x}$ and $\bar{A}$ correspond to the fixed part

$$\min_{x_i} \quad \frac{1}{2}(A_i x_i + \bar{A}\bar{x} - b)^2 + \lambda|x_i|$$

Soft-thresholding operator

$$\min_{x_i} \frac{1}{2}(x_i - r)^2 + \lambda|x| \rightarrow x_i = \begin{cases} r - \lambda & \text{if } r > \lambda \\ 0 & \text{if } -\lambda \leq r \leq \lambda \\ r + \lambda & \text{if } r < -\lambda \end{cases}$$

$$r = -A_i{}^\top(\bar{A}\bar{x} - b)/||A_i||^2, \; \lambda \rightarrow \lambda/||A_i||^2$$

Remember coordinate descent for Lasso

$$\min_{x_i} \quad \frac{1}{2}\|W_{11}^{1/2}\beta - W_{11}^{-1/2}a_{12}\|^2 + \lambda\|\beta\|_1$$

$$\min_{\beta_i} \frac{1}{2}(\beta_i - r)^2 + \lambda|x| \rightarrow \beta_i = \begin{cases} r - \lambda & \text{if } r > \lambda \\ 0 & \text{if } -\lambda \le r \le \lambda \\ r + \lambda & \text{if } r < -\lambda \end{cases}$$

$$r = -((W_{11})_i{}^\top \bar{\beta} - (a_{12})_i)/(W_{11})_{ii}, \;\; \lambda \rightarrow \lambda/(W_{11})_{ii}$$

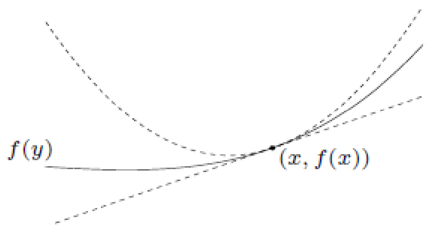
No need to compute $W^{1/2}$

Prox method with nonsmooth term

- Consider: $$\min_x F(x) = f(x) + g(x)$$

$$|\nabla f(x) - \nabla f(y)| \leq L||x - y||$$

- Quadratic upper approximation



$$f(y) + g(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2\mu}||y - x||^2 + g(y) = Q_{f,\mu}(x, y)$$

$$F(y) \leq f(x) + \frac{1}{2\mu}||x - \mu \nabla f(x)^\top - y||^2 + g(y) = Q_{f,\mu}(x, y)$$

Assume that *g(y)* is such that the above function is easy to optimize over *y*

ISTA/Gradient prox method

$$\min_x F(x) = f(x) + g(x)$$

- Minimize quadratic upper approximation on each iteration

$$x^{k+1} = \text{argmin}_y Q_f(x^k, y)$$

$$Q_{f,\mu}(x, y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2\mu}||y - x||^2 + g(y)$$

- If $\mu \leq 1/L$ then in $O(L/\epsilon)$ iterations finds solution

$$\bar{x} : \; F(\bar{x}) \leq F(x^*) + \epsilon$$

Fast first-order method

Nesterov, Beck & Teboulle

$$\min_x F(x) = f(x) + g(x)$$

- Minimize upper approximation at an "accelerated" point.

$$x^k = \text{argmin}_y Q_f(y^k, y)$$

$$t_{k+1} := \quad (1 + \sqrt{1 + 4t_k^2})/2$$

$$y^{k+1} := \quad x^k + \frac{t_k - 1}{t_{k+1}}[x^k - x^{k-1}]$$

- If $\mu \leq$ *1/L* then in $O(\sqrt{L/\epsilon})$ iterations finds solution

$$\bar{x} : \; F(\bar{x}) \leq F(x^*) + \epsilon$$

Example 1 (Lasso and SICS)

$$\min_x f(x) + ||x||_1$$

- Minimize upper approximation function $Q_{f,\mu}(x,y)$ on each iteration

$$\min_y Q_{f,\mu}(x,y) = \min_y f(x) + \frac{1}{2\mu}||x - \mu\nabla f(x)^\top - y||^2 + ||y||_1$$

$$\Updownarrow$$

$$\sum_i \min_{y_i}\left[\frac{1}{2\mu}(y_i - r_i)^2 + |y_i|\right]$$

$$\Updownarrow$$

$$\min_{y_i} \frac{1}{2}(y_i - r_i)^2 + \mu|y_i| \rightarrow y_i^* = \begin{cases} r_i - \mu & \text{if } r_i > \mu \\ 0 & \text{if } -\lambda \leq r_i \leq \mu \\ r_i + \mu & \text{if } r_i < -\mu \end{cases}$$

Closed form solution! *O(n)* effort

First order method

$$\max_{C\succ 0} \frac{m}{2}(\ln\det(C) - Tr(AC)) - \lambda\|C\|_1$$

Can be written in the form $\max_{C\succ 0} F(C) - \lambda\|C\|_1$

Given C:

$$\begin{aligned} C^+ &= \text{argmax}_X F(C) + < X - C, \nabla F(C) > \\ &+ \frac{1}{2\mu}\|C - X\|_F^2 - \lambda\|X\|_1 \end{aligned}$$

The step of a first order method
with gradient given below

$$\nabla F(C) = C^{-1} - A$$

First order method

$$
\begin{aligned}
C^{+} &= \mathrm{argmax}_{\mathrm{X}} F(C) + < X - C, \nabla F(C) > \\
&+ \frac{1}{2\mu}\|C - X\|_F^2 - \lambda\|X\|_1
\end{aligned}
$$

$$\nabla F(C) = C^{-1} - A$$

Each step requires gradient computation $O(n^3)$

Fails if C^{+} is not psd.

Collaborative prediction

$$\min_{X \in \mathrm{R}^{n \times m}} \lambda \sum_{(i,j) \in I} (X_{ij} - M_{ij})^2 + ||X||_*$$

Find matrix X of the lowest rank that has given entries

$$\min_{X \in \mathrm{R}^{n \times m}} \quad rank(X) = \|\sigma_i(X)\|_0$$

$$\text{s.t.} \quad X_{ij} = M_{ij}, \ (i,j) \in I$$

Nuclear norm of matrix *X*

$$\|X\|_* = \sum_i \sigma_i(x)$$

is the tightest convex relaxation of rank(*X*)

movies

users

	2		1			4				5	
	5		4				?		1		3
		3		5			2				
4			?			5		3		?	
		4		1	3				5		
			2				1	?			4
	1					5		5		4	
		2		?	5		?		4		
	3		3		1		5		2		1
	3				1			2		3	
	4			5	1			3			
		3				3	?			5	
2	?		1		1						
		5			2	?		4		4	
	1		3		1	5		4		5	
1		2			4				5	?	

Example 3 (Collaborative Prediction)

$$\min_{X \in \mathrm{R}^{n \times m}} f(X) + ||X||_*$$

$$\min_Y Q_f(X, Y)$$

$$\Updownarrow$$

$$\min_Y \left[\frac{1}{2\mu} \|Y - Z\|_F^2 + \|Y\|_* \right]$$

$$\Updownarrow$$

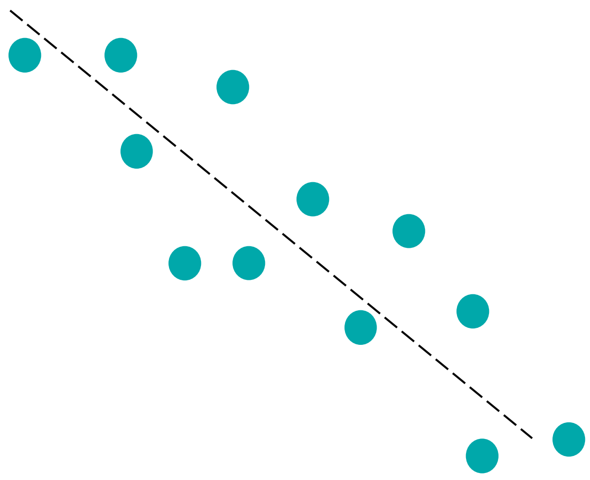
$$Z = P\mathrm{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_n\} Q^\top$$

Closed form solution! *O(n^3)* effort

$$Y^* = P\mathrm{diag}\{\sigma_1^*, \sigma_2^*, \ldots, \sigma_n^*\} Q^\top, \; \sigma_i^* = \begin{cases} \sigma_i - \mu & \text{if } \sigma_i > \mu \\ 0 & \text{if } -\mu \le \sigma_i \le \mu \\ \sigma_i + \mu & \text{if } \sigma_i < -\mu \end{cases}$$

Group Lasso regression

- Problem: $$\min_x \quad \frac{1}{2}||Ax-b||^2 + \lambda \sum_i ||x_i||_2$$



• Assume that columns of A form groups of correlated features.
• Find sparse vector x where nonzeros are selected according to groups
• x_i is a subvector of x corresponding to the i-th group of features.

Example 2 (Group Lasso)

$$\min_x f(x) + \sum_i ||x_i||, \;\; x_i \in \mathrm{R}^{n_i}$$

- Very similar to the previous case, but with ||.|| instead of |.|

$$\sum_i \min_{y_i \in \mathrm{R}^{n_i}} \left[\frac{1}{2\mu}(y_i - r_i)^2 + \|y_i\| \right]$$

$$\Updownarrow$$

$$y_i^* = \frac{r_i}{\|r_i\|} \max(0, \|r_i\| - \mu)$$

Closed form solution! *O(n)* effort