CS369M: Algorithms for Modern Massive Data Set Analysis **Lecture 18, - 12/02/2009**

Data-motivated Matrix Factorizations (2 of 2)

Lecturer: Michael Mahoney *Scribe: Bahman Bahmani*

**Unedited Notes*

1 Rank Minimization

The general rank minimization problem, which arises in a wide range of applications, is as follows:

$$\min. \quad rank(X) \tag{1}$$
$$\text{s.t.} \quad X \in \mathcal{C} \tag{2}$$

where $\mathcal{C}$ is a convex subset of $R^{m\times n}$. Since this problem is generally hard to solve, we replace it with the following intuitively sound optimization:

$$\min. \quad ||X||_\star \tag{3}$$
$$\text{s.t.} \quad X \in \mathcal{C} \tag{4}$$

where $||X||_\star = \sum_i \sigma_i(X)$ is the sum of singular values of X.

Even though the original rank-minimization problem is non-convex, the above heuristic optimization is indeed convex. Also, we have the following theorem, which shows this is actually a good convex formulation:

Theorem 1 *$||X||_\star$ is the convex envelope of $rank(X)$ on $\{X \in R^{m\times n} | \ ||X|| \leq 1\}$.*

The proof of this theorem can be found in [1].

As mentioned, the heuristic formulation is a convex problem, and hence can be solved in general. We also show that for the special case where $\mathcal{C}$ is a set of linear constraints, we can turn this problem into an SDP. The problem is equivalent to:

$$\min. \quad t \tag{5}$$
$$\text{s.t.} \quad ||X||_\star \leq t \tag{6}$$
$$X \in \mathcal{C} \tag{7}$$

But, we have the following lemma:

Lemma 2 *For $X \in R^{m\times n}$ and $t \in R$, $||X||_\star \leq t$ iff there exist matrices $Y \in R^{m\times m}$ and $Z \in R^{n\times n}$ such that:*

$$\begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0, \qquad tr(Y) + tr(Z) \leq 2t$$

Hence, the last optimization is equivalent to:

$$\begin{aligned} \text{min.} \quad & tr(Y) + tr(Z) & (8) \\ \text{s.t.} \quad & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 & (9) \\ & X \in \mathcal{C} & (10) \end{aligned}$$

which is an SDP (if $\mathcal{C}$ is a set of linear constraints) and hence can be solved efficiently using any SDP solver.

2 Maximum Margin Matrix Factorization

Assume we have a matrix $Y \in \{\pm 1\}^{n \times m}$ some subset S of whose entries have been observed (and formed Y_S). We would like to approximate the rest of the entries. To do so, we can find an approximation X of Y using an optimization over the observed entries. One way to do so is to find a low-rank approximation X. Notice that $rank(X) \leq k$ iff X can be written as UV^T where $U \in R^{n \times k}$ and $V \in R^{m \times k}$. Hence, looking for low rank X corresponds to seeking low dimensionality factorization.

Another approach is looking for small norm factorization (through a penalty term), where norm of the factorization is measured by $||U||_{Fro}^2 + ||V||_{Fro}^2$. We have the following lemma [2]:

Lemma 3

$$min_{X=UV^T} \frac{1}{2}(||U||_{Fro}^2 + ||V||_{Fro}^2) = min_{X=UV^T} ||U||_{Fro} ||V||_{Fro} = ||X||_\star$$

Hence, using the above approach and the above lemma, we can formulate two optimization variants:

1. Hard-margin matrix factorization

$$\begin{aligned} \text{min.} \quad & ||X||_\star & (11) \\ \text{s.t.} \quad & Y_{ia} X_{ia} \geq 1 \quad \forall\, ia \in S & (12) \end{aligned}$$

2. Soft-margin matrix factorization

$$\text{min.} ||X||_\star + c \sum_{ia \in S} max(0, 1 - Y_{ia} X_{ia})$$

Now, using lemma 2, we can write the soft-margin optimizations as follows:

$$\begin{aligned} \text{min.} \quad & \frac{1}{2}(tr(A) + tr(B)) + c \sum_{ia \in S} \xi_{ia} & (13) \\ \text{s.t.} \quad & \begin{bmatrix} A & X \\ X^T & B \end{bmatrix} \geq 0 & (14) \\ & y_{ia} X_{ia} \geq 1 - \xi_{ia} \quad \forall\, ia \in S & (15) \\ & \xi_{ia} \geq 0 \quad \forall\, ia \in S & (16) \end{aligned}$$

The hard-margin optimization can also be written similarly (with slack variables equal to zero). This is an SDP and hence can be solved efficiently.

References

[1] Fazel, Hindi, and Boyd, "A Rank Minimization Heuristic with Application to Minimum Order System Approximation"

[2] Srebro, Rennie, and Jaakkola, "Maximum Margin Matrix Factorizations"