低秩矩阵恢复

- 某视频网站提供了约48万用户对1万7千多部电影的上亿条评级数据，希望对用户的电影评级进行预测，从而改进用户电影推荐系统，为每个用户更有针对性地推荐影片．
- 显然每一个用户不可能看过所有的电影，每一部电影也不可能收集到全部用户的评级．电影评级由用户打分1 星到5 星表示，记为取值1~5 的整数．我们将电影评级放在一个矩阵M中，矩阵M的每一行表示不同用户，每一列表示不同电影．由于用户只对看过的电影给出自己的评价，矩阵M中很多元素是未知的

$$
\begin{array}{c}
 \\ \text{用户1} \\ \text{用户2} \\ \text{用户3} \\ \text{用户4} \\ \vdots \\ \text{用户m}
\end{array}
\begin{array}{c}
\begin{array}{cccccc} \text{电影1} & \text{电影2} & \text{电影3} & \text{电影4} & \dots & \text{电影n} \end{array} \\
\begin{bmatrix}
4 & ? & ? & 3 & \cdots & ? \\
? & 2 & 4 & ? & \cdots & ? \\
3 & ? & ? & ? & \cdots & ? \\
2 & ? & 5 & ? & \cdots & ? \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
? & 3 & ? & 4 & \cdots & ?
\end{bmatrix}
\end{array}
$$

低秩矩阵恢复问题的性质

该问题在推荐系统、图像处理等方面有着广泛的应用。

- 由于用户对电影的偏好可进行分类，按年龄可分为：年轻人，中年人，老年人；且电影也能分为不同的题材：战争片，悬疑片，言情片等。故这类问题隐含的假设为补全后的矩阵应为低秩的。即矩阵的行与列会有“合作”的特性，故该问题具有别名“collaborative filtering”。

- 除此之外，由于低秩矩阵可分解为两个低秩矩阵的乘积，所以低秩限制下的矩阵补全问题是比较实用的，这样利于储存且有更好的诠释性。

- 有些用户的打分可能不为自身真实情况，对评分矩阵有影响，所以原矩阵是可能有噪声的。

低秩矩阵恢复

由上述分析可以引出该问题：

- 令Ω 是矩阵M 中所有已知评级元素的下标的集合，则该问题可以初步描述为构造一个矩阵X，使得在给定位置的元素等于已知评级元素，即满足$X_{ij} = M_{ij}, \ (i,j) \in \Omega$·
- 低秩矩阵恢复（low rank matrix completion）

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \quad & \operatorname{rank}(X), \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \ (i,j) \in \Omega. \end{aligned} \tag{6}$$

rank(X)正好是矩阵X所有非零奇异值的个数

- 矩阵X的核范数（nuclear norm）为矩阵所有奇异值的和，即：$\|X\|_* = \sum_i \sigma_i(X)$:

$$\begin{aligned} \min_{X \in \mathbb{R}^{m \times n}} \quad & \|X\|_*, \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \ (i,j) \in \Omega. \end{aligned} \tag{7}$$

- 可以证明问题(7) 是一个凸优化问题，并且在一定条件下它与问题(6) 等价.
- 也可以将问题(7) 转换为一个半定规划问题，但是目前半定规划算法所能有效求解的问题规模限制了这种技术的实际应用.
- 考虑到观测可能出现误差，对于给定的参数$\mu > 0$，给出该问题的二次罚函数形式：

$$\min_{X \in \mathbb{R}^{m \times n}} \quad \mu \|X\|_* + \frac{1}{2} \sum_{(i,j) \in \Omega} (X_{ij} - M_{ij})^2. \tag{8}$$

低秩矩阵恢复

- 秩r情形：$X = LR^T$,其中$L \in \mathbb{R}^{m\times r}$,$R \in \mathbb{R}^{n\times r}$ 并且$r \ll \min(m,n)$.则可将问题写为

$$\min_{\boldsymbol{L},\boldsymbol{R}} \sum_{(i,j)\in\omega} \left(\left[\boldsymbol{L}\boldsymbol{R}^T\right]_{ij} - M_{ij} \right)^2 + \alpha \|\boldsymbol{L}\|_F^2 + \beta \|\boldsymbol{R}\|_F^2$$

- 在该问题中，矩阵X在定义中已为秩r矩阵，所以没有必要再加上秩约束正则项。α, β为正则化参数，这里正则化的作用是消除解L, R在放缩意义下的不唯一性.
- 此时L, R矩阵中的数字之和为$(m+n)r$,远小于np,不过此时问题是非凸的.
- 尽管这个该问题是非凸的，但在某种意义上它是一个可处理问题的近似：如果对X 有一个完整的观察，那么秩-r近似可以通过X的奇异值分解来找到，并根据r导出的左奇异向量和右奇异向量定义L和R。