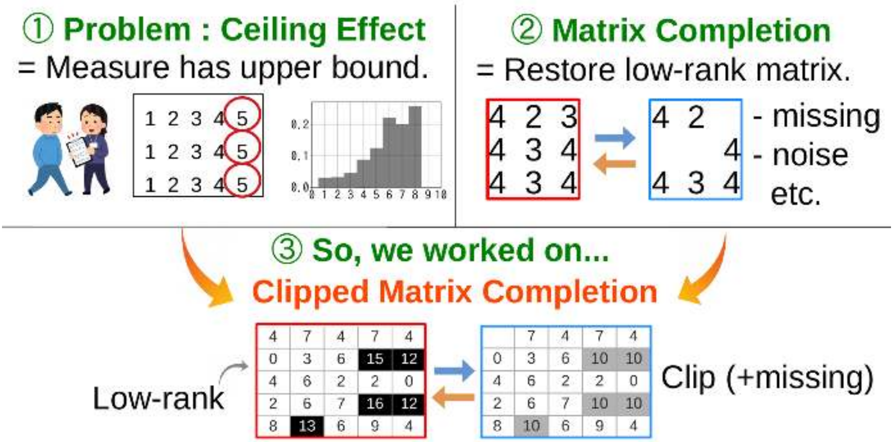
Overview


① Problem : Ceiling Effect
= Measure has upper bound.
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
0.2
0.1
0.0
0 1 2 3 4 5 6 7 8 9 10
② Matrix Completion
= Restore low-rank matrix.
4 2 3
4 3 4
4 3 4
4 2
4
4 3 4
- missing
- noise
etc.
③ So, we worked on...
Clipped Matrix Completion
Low-rank
4 7 4 7 4
0 3 6 15 12
4 6 2 2 0
2 6 7 16 12
8 13 6 9 4
7 4 7 4
0 3 6 10 10
4 6 2 2 0
2 6 7 10 10
8 10 6 9 4
Clip (+missing)

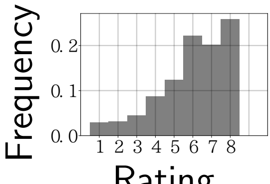
①Ceiling effect（Important!）

Ceiling Effect

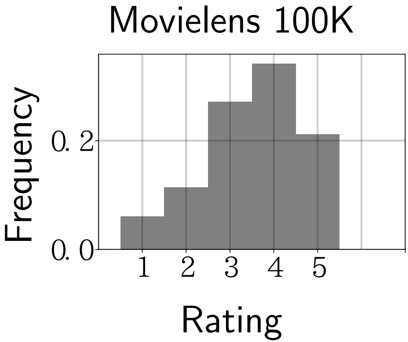
Situation: measurement tool has an upper bound [1]. Exceeding values are observed after clipping to the upper bound.

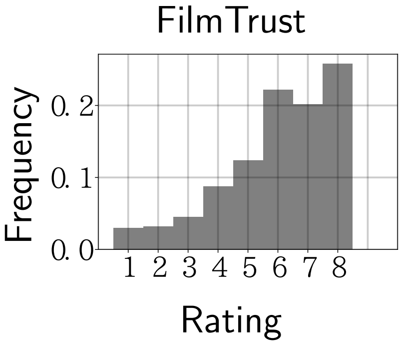
- Ex. 5-scale response

- Too many “5” ⇒ question failed to capture true preferences.



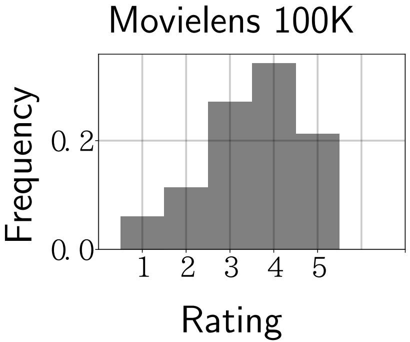
Ceiling effect in machine learning data

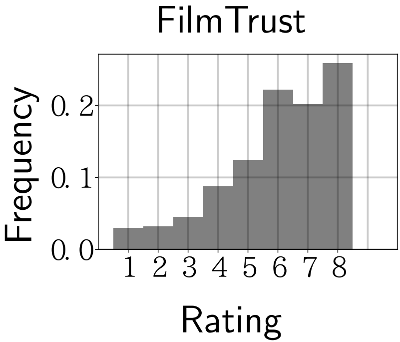




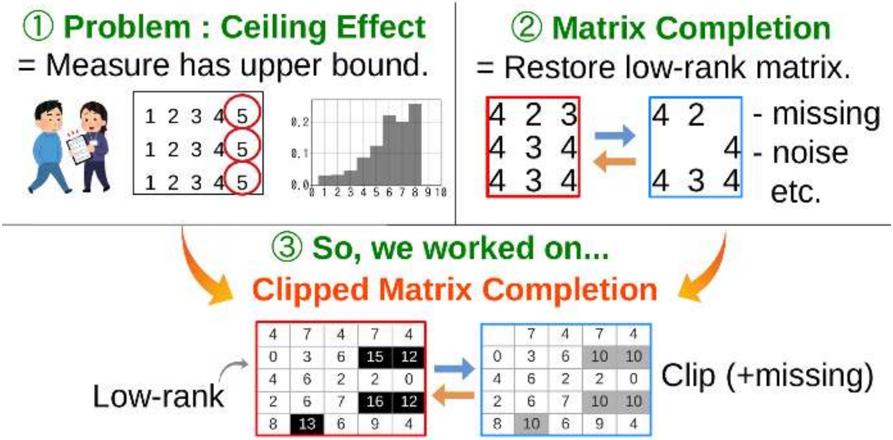
- Benchmark data of recommendation systems.
- Right-truncated shape is typical for ceiling effect

How can we investigate the true values of data that is prone to ceiling effects?



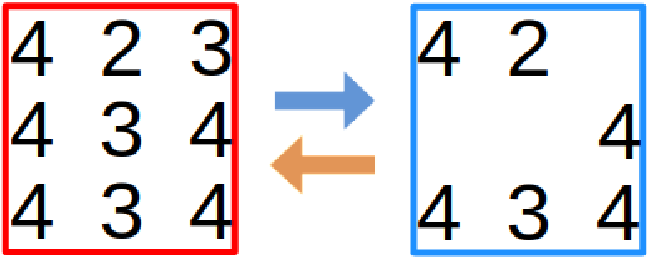


Overview


① Problem : Ceiling Effect
= Measure has upper bound.
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
② Matrix Completion
= Restore low-rank matrix.
4 2 3
4 3 4
4 3 4
4 2
4
4 3 4
- missing
- noise
etc.
③ So, we worked on...
Clipped Matrix Completion
Low-rank
4 7 4 7 4
0 3 6 15 12
4 6 2 2 0
2 6 7 16 12
8 13 6 9 4
7 4 7 4
0 3 6 10 10
4 6 2 2 0
2 6 7 10 10
8 10 6 9 4
Clip (+missing)

②Matrix completion

- A technique to recover a matrix from deficits e.g. missing [3].



- Goal: fill in the blanks
- Missing, noise, quantization, etc. For each deficit, methods are developed.

- Example application: movie recommendation system

Movie

User

	5		4
5		1	5
5	1	3	3

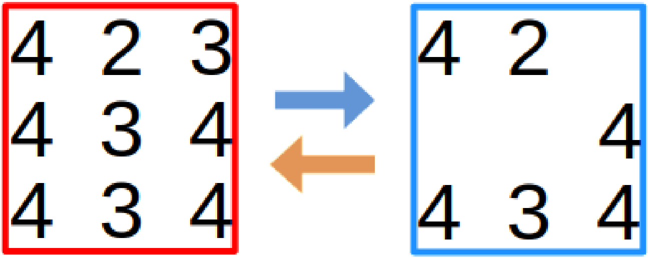
②How matrix completion works

- There are conditions for MC to be possible.
- Low-rank and well-behaved matrix ⇒ recoverable [3].

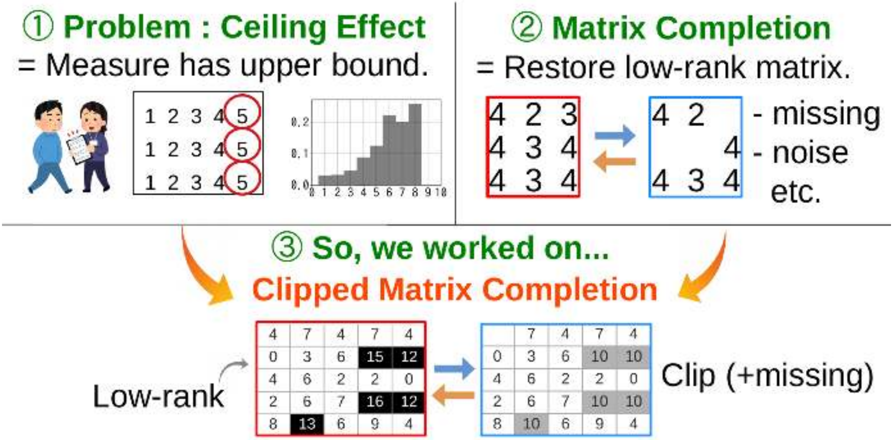
The principle of low-rank completion

- Low-rank ⋯ each entry is an inner product of row-/column-vectors.
- How to fill in: estimate feature vectors → compute inner products

The principle of low-rank matrix completion: recover matrices from missing etc.



Overview


① Problem : Ceiling Effect
= Measure has upper bound.
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
0.2
0.1
0.0
0 1 2 3 4 5 6 7 8 9 10
② Matrix Completion
= Restore low-rank matrix.
4 2 3
4 3 4
4 3 4
4 2
4
4 3 4
- missing
- noise
etc.
③ So, we worked on...
Clipped Matrix Completion
Low-rank
4 7 4 7 4
0 3 6 15 12
4 6 2 2 0
2 6 7 16 12
8 13 6 9 4
7 4 7 4
0 3 6 10 10
4 6 2 2 0
2 6 7 10 10
8 10 6 9 4
Clip (+missing)

Completing low-rank matrix from its clipped observations

	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Observation

⇒

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

Underlying matrix

Problem (Clipped Matrix Completion; CMC)

Accurately recover the underlying matrix from a random subset of its clipped observations and the known clipping threshold.

List of results and contributions

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

True values (unknown)

Obs. ⇒

	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Observed data (clip @ 10)

Rec. ⇒

4.0	7.0	4.0	7.0	4.0
-0.0	3.0	6.0	14.9	11.9
4.0	6.0	2.0	2.0	0.0
2.0	6.0	7.0	15.9	11.9
8.0	13.0	6.0	9.0	4.0

After completion

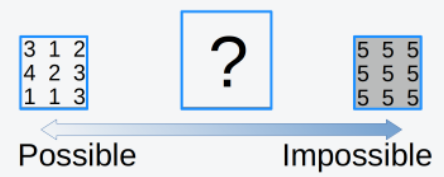
1. Proposed problem setting → Clipped matrix completion.
2. Recovery possible? → Exact recovery possible under conditions.
3. Recovery method? → Minimize squared hinge loss + regularization term
4. Experimentally? → Resilience to ceiling effect may improve recommendation systems!

List of technical details

1. Main theorem: when is recovery possible?
2. Proposed method: how to do the recovery?
3. Theoretical guarantee of the proposed method (omitted)
4. Experimental evaluation

Technical detail ①: when is recovery possible?

- Motivation for theory: not all clipped matrices can be completed.



  - There are evident cases where recovery is impossible.
  - There are cases where no treatment for clipping is required.

- Main theorem: a sufficient condition for the recovery to be possible.
  → at least, there are cases where recovery is feasible (even with non-negligible clipping).

Technical detail ①Main theorem

Assumptions (informal)

1. The effect of clipping (definition is involved) is small.
2. True matrix M is low-rank.
3. M is "Incoherent"(a small subset of observation is sufficient for estimating the entire matrix)
4. The elements are observed independently with high-enough probability p.

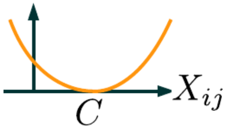
Theorem (Exact recovery for CMC; informal)

With high probability, with a certain algorithm (trace-norm minimization), the true matrix can be recovered exactly.

Technical detail ②: how to recover?

- Ordinary MC [5] : square loss

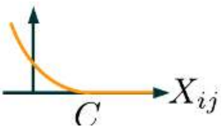
$$\arg\min_{\mathbf{X}} \frac{1}{2} \sum_{ij\,:\,\mathrm{obs}} (\mathrm{Obs}_{ij} - X_{ij})^2 + \mathcal{R}(\mathbf{X})$$



- On the clipped entries, C is wrongly recovered.

- CMC (proposed): square hinge loss

$$\arg\min_{\mathbf{X}} \frac{1}{2} \sum_{ij\,:\,\text{non-clip. obs.}} (\mathrm{Obs}_{ij} - X_{ij})^2 + \frac{1}{2} \sum_{ij\,:\,\text{clipped obs.}} \max(0, \mathrm{Obs}_{ij} - X_{ij})^2 + \mathcal{R}(\mathbf{X})$$



Design of regularization term: induce low-rank solution

1. DTr-CMC: Double trace-norm regularization (proposed)
   - Effect: induce low-rankness in both $\mathbf{X}$ and $\mathrm{Clip}(\mathbf{X})$
   - Theoretical guarantee is also given (details omitted)
2. Tr-CMC: trace-norm regularization [5]
   - Effect: induce low-rank solution
3. Fro-CMC: Frobenius norm regularization [4]
   - Effect: induce low-rank solution

Details of the regularization terms

1. DTr-CMC: Double trace-norm regularization

   $\mathcal{R}(\mathbf{X}) = \lambda_1 \|\mathbf{X}\|_{\mathrm{tr}} + \lambda_2 \|\mathrm{Clip}(\mathbf{X})\|_{\mathrm{tr}}$ $\mathrm{Clip} = \min(\cdot, C)$

   - Effect: induce low-rankness in both $\mathbf{X}$ and $\mathrm{Clip}(\mathbf{X})$
   - Optimization: (approximate) subgradient descent [2]
   - Theoretical guarantee is also given (details omitted)

2. Tr-CMC: trace-norm regularization [5]

   $\mathcal{R}(\mathbf{X}) := \lambda \|\mathbf{X}\|_{\mathrm{tr}}$ $\|\mathbf{X}\|_{\mathrm{tr}} = \sum_{l=1}^{\min(n_1,n_2)} \sigma_l$

   (σ_l: l-th singular value)

   - Effect: induce low-rank solution
   - Optimization: accelerated gradient descent [5]

3. Fro-CMC: Frobenius norm regularization [4]

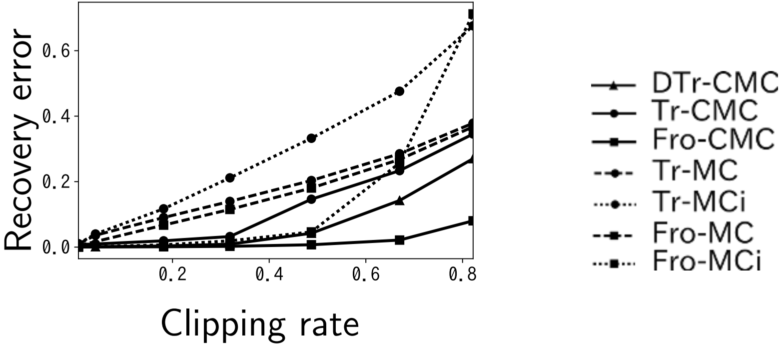
   $\mathcal{R}(\mathbf{P}, \mathbf{Q}) := \lambda_1 \|\mathbf{P}\|_{\mathrm{F}}^2 + \lambda_2 \|\mathbf{Q}\|_{\mathrm{F}}^2$ $\mathbf{X} = \mathbf{P}\mathbf{Q}^\top$

   - Effect: induce low-rank solution
   - Optimization: (approximate) alternating least squares [4]

Technical details ③: experiments

1. Experiment with synthetic data
   - Evaluate the recovery result under a controlled situation (where true value is known)
2. Experiment with real-world data
   - The true values are unknown $\Rightarrow$ evaluation of recovery is impossible.
     (The true values before clipping in real-world data is unknown)
   - Device for evaluation: evaluate a binary (two-class) classification task to classify entries into "the true value is above threshold or not."

Experiment 1/3 Experiment with synthetic data



- Solid: proposed method, dotted: baseline methods.
- Vary clipping threshold $\rightarrow$ eval. test recovery error.
- Proposed method (solid) is able to estimate the true matrix with small error of order 10^{-2} even when there is 70% clipping.

Experiment 2/3 Experiment with real-world data (1)

f_1 value	DTr-CMC	Fro-CMC	Fro-MC	Tr-CMC	Tr-MC	(baseline)
Film Trust	0.47 (0.01)	0.35 (0.01)	0.27 (0.01)	0.36 (0.00)	0.22 (0.00)	0.41 (0.00)
Movielens 100K	0.39 (0.00)	0.41 (0.00)	0.21 (0.01)	0.40 (0.00)	0.12 (0.00)	0.35 (0.00)

- Learning after artificially clipping real data (★ 5 →★ 4)
- Classify true ratings into "★ 5 $\geq$" and "$\leq$ ★ 4"
- Proposed method estimates the true value better
- (Baseline: a classifier which unconditionally outputs $+1$)

Experiment 3/3 Experiment with real-world data (2)

f_1 value	DTr-CMC	Fro-CMC	Fro-MC	Tr-CMC	Tr-MC	(baseline)
Film Trust	0.46 (0.01)	0.40 (0.01)	0.35 (0.01)	0.39 (0.00)	0.35 (0.01)	0.41 (0.00)
Movielens 100K	0.38 (0.00)	0.41 (0.01)	0.38 (0.01)	0.40 (0.00)	0.38 (0.00)	0.35 (0.00)

- Learning from real data (without artificial clipping) (★ 1∼★ 5)
- Classify true ratings into "★ 5 $\geq$" and "$\leq$ ★ 4"
- Robustness to ceiling effect improves the detection power of high-rating entries.
- (Baseline: a classifier which unconditionally outputs $+1$)

Summary

- Problem setting? → Recover matrix from ceiling effects
- Recovery possible? → Exact recovery possible under conditions
- How to recover? → Minimize square hinge loss + regularization term
- Experimentally? → Resilience to ceiling effect may improve recommendation systems!

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

True values (unknown)

Obs. ⇒

	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

Observed data (clip @ 10)

Rec. ⇒

4.0	7.0	4.0	7.0	4.0
-0.0	3.0	6.0	14.9	11.9
4.0	6.0	2.0	2.0	0.0
2.0	6.0	7.0	15.9	11.9
8.0	13.0	6.0	9.0	4.0

After completion

Appendix

Coherence

- Let $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ (Singular value decomposition).
- Define *coherence* by $\mu_0 := \max\left\{\frac{n_1}{r}\mu^{\mathrm{U}}(\mathbf{M}), \frac{n_2}{r}\mu^{\mathrm{V}}(\mathbf{M})\right\}$
  - Here, $\mu^{\mathrm{U}}(\mathbf{M}) := \max_{i\in[n_1]} \|\mathbf{U}_{i,\cdot}\|^2$, $\mu^{\mathrm{V}}(\mathbf{M}) := \max_{j\in[n_2]} \|\mathbf{V}_{j,\cdot}\|^2$, $r = \mathrm{rank}(\mathbf{M})$.
- Joint coherence is defined by $\mu_1 := \sqrt{\frac{n_1 n_2}{r}}\|\mathbf{U}\mathbf{V}^\top\|_\infty$
- $\mathbf{M}$ is *incoherent* if both μ_0 and μ_1 are small.

Information deficit

- $\mathcal{B} := \{(i,j) : M_{ij} < C\}$
- $$T := \operatorname{span}\left(\{\boldsymbol{u}_k \boldsymbol{y}^\top : k \in [r], \boldsymbol{y} \in \mathbb{R}^{n_2}\} \cup \{\boldsymbol{x}\boldsymbol{v}_k^\top : k \in [r], \boldsymbol{x} \in \mathbb{R}^{n_1}\}\right)$$
- $(\mathcal{P}^*(\mathbf{Z}))_{ij} := \mathbf{1}\{M_{ij} < C\} Z_{ij} + \mathbf{1}\{M_{ij} = C\}(Z_{ij})_+$
- $\rho_{\mathrm{F}} := \sup_{\mathbf{Z} \in T \setminus \{\mathbf{O}\} : \|\mathbf{Z}\|_{\mathrm{F}} \le \|\mathbf{U}\mathbf{V}^\top\|_{\mathrm{F}}} \frac{\|\mathcal{P}_T \mathcal{P}^*(\mathbf{Z}) - \mathbf{Z}\|_{\mathrm{F}}}{\|\mathbf{Z}\|_{\mathrm{F}}}$
- $\rho_{\infty} := \sup_{\mathbf{Z} \in T \setminus \{\mathbf{O}\} : \|\mathbf{Z}\|_{\infty} \le \|\mathbf{U}\mathbf{V}^\top\|_{\infty}} \frac{\|\mathcal{P}_T \mathcal{P}^*(\mathbf{Z}) - \mathbf{Z}\|_{\infty}}{\|\mathbf{Z}\|_{\infty}}$
- $\rho_{\mathrm{op}} := \sqrt{r}\mu_1 \left(\sup_{\substack{\mathbf{Z} \in T \setminus \{\mathbf{O}\} : \\ \|\mathbf{Z}\|_{\mathrm{op}} \le \sqrt{n_1 n_2} \|\mathbf{U}\mathbf{V}^\top\|_{\mathrm{op}}}} \frac{\|\mathcal{P}^*(\mathbf{Z}) - \mathbf{Z}\|_{\mathrm{op}}}{\|\mathbf{Z}\|_{\mathrm{op}}} \right)$
- $\nu_{\mathcal{B}} := \|\mathcal{P}_T \mathcal{P}_{\mathcal{B}} \mathcal{P}_T - \mathcal{P}_T\|_{\mathrm{op}}$

Can we account for floor effect in addition to ceiling effect?

- Yes, the result is extendable to floor effect (cf. the paper).
- Element-wise thresholds are also allowed.

How to determine if recovery is possible

- The assumptions can be checked only after seeing the true matrix.
- However, there are some intuitions what kind of matrix is possible to recover.
  - Low-rank = consists of a small number of components
  - A few (column-/row-wise) common factors almost determine the matrix entries.
  - In other words, there are similar rows/columns.
  - The space spanned by the singular vectors of $\mathbf{M}$ is not aligned with the indicator-matrices of the indices which are to be clipped.
    - Roughly speaking, rank-one matrices of SVD (the components of $\mathbf{M}$) have support all over the indices (if there are sparse components which may have large values on clipped indices, the recovery is impossible).

How to determine hyper-parameters?

- No theoretically justified method for hyper-parameter selection in recovery problems.
  - In synthetic data experiment, we selected the parameter with the smallest difference between the data and the clipped version of the estimated matrix.
- In the real data experiment, the final performance can be computed. Therefore, we used the one with the best performance on a held-out validation indices.
  - Similarly, in recommendation systems, the final performance measure is likely available for hyper-parameter selection.

Future work for clipped matrix completion

- Characterize necessary condition for recovery.
- Develop algorithms to perform artificial clipping to disable a recovery by arbitrary method.

Trace-norm minimization

Trace-norm minimization is the algorithm defined as below.

$$\arg\min_{\mathbf{X}} \|\mathbf{X}\|_{\mathrm{tr}} \text{ s.t. } \begin{cases} \mathcal{P}_{\Omega\setminus\mathcal{C}}(\mathbf{X}) = \mathcal{P}_{\Omega\setminus\mathcal{C}}(\mathbf{M}^{\mathrm{c}}_{\Omega}), \\ \mathcal{P}_{\mathcal{C}}(\mathbf{M}^{\mathrm{c}}_{\Omega}) \leq \mathcal{P}_{\mathcal{C}}(\mathbf{X}), \end{cases}$$

Here, $\Omega := \{(i,j) : \text{observed}\}$ and $\mathcal{C} := \{(i,j) \in \Omega : M^{\mathrm{c}}_{ij} = C\}$.

Details of the real-world data experiment

- $\mathrm{precision}$: Among those predicted "yes," the fraction of true "yes." (ratio of precise predictions)
- recall : Among those with true "yes," the fraction of those predicted "yes." (ratio of correctly recalled true "yes")
- $\mathrm{f}_1 = \frac{2 \cdot \mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$
- f_1 was used because in this binary classification, the challenge is to extrapolate to a large value from observed small values. Therefore, recall is considered as the difficult part.

Reference

[1] Encyclopedia of research design.

[2] Haim Avron, Satyen Kale, Shiva Prasad Kasiviswanathan, and Vikas Sindhwani.
Efficient and practical stochastic subgradient descent for nuclear norm regularization.
In *Proceedings of the 29th International Conference on Machine Learning*, pages 1231–1238.

[3] Emmanuel J. Cands and Yaniv Plan.
Matrix completion with noise.
98(6):925–936.

[4] Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi.
Low-rank matrix completion using alternating minimization.
In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pages 665–674.

[5] Kim-Chuan Toh and Sangwoon Yun.
An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems.
6:615–640.