Matrix completion

-

IISc

May 2,2012

Degrees of freedom in a matrix(n×n) of rank r

r(2n-r)

Proof of $r(2n-r)$

$\Rightarrow nr + (n-r)r$
$\Rightarrow nr + nr - r^2$
$\Rightarrow 2nr - r^2$

- This will be of Θ(nr) for a n×n matrix of rank r.
- which is much smaller than $\Theta(n^2)$.

Example practical situations of low ranks

For a matrix 1000x1000 of rank 20 we need just 40000 numbers instead of 1000000 numbers.

Which set of $r(2n-r)$ numbers?

Ofcourse not all sets of such numbers will work.

1. If a complete row or column is missing, then you can't recover. There is no hope. how there is no hope
2. If the matrix is something like this.

$$M = e_1 e_n^* = \begin{bmatrix} 0 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

3. or

$$M = e_1 x^* = \begin{bmatrix} x_1 & \cdots & x_n \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

explain how? by comparing with another matrix

4. So there are bunch of cases in which matrix recovery isn't going to work, whatsoever the method is..

Sample the matrix at random?

- Only some combinitions of $r(2n-r)$ numbers are going to be the right choice. (Good cases/ Bad cases)
  - Matrix.
  - No of enteries. $(> r(2n-r))$
  - Which combination of these enteries.

- What is **Probability** of exact matrix completion. Randomly sample
  1. Matrix - Number of combinations of numbers that are GOOD cases.
  2. Number of samples $(> r(2n-r)$, otherwise its 0)

Central Idea!!

If I want the matrix to get completed for **most of cases (fixed probability)**, we will need **different amount of enteries** $\mathbf{(} > r(2n-r)\mathbf{)}$ for different matrices.

explain P=m. x(M)

$$Samlpes \propto \Psi_X(M) \text{ given some P}$$

Coherence

$$L \in R^{n \times n} = U \Sigma V^*$$

Coherence parameter $\mu \geq 1$

$\| P_U e_i \|^2 \leq \frac{\mu r}{n}$ $\quad \| P_V e_i \|^2 \leq \frac{\mu r}{n}$ where

- P_i is projection onto corrosponding space.
- e_i is i_{th} unit vector.

Alternatively

$$\mu(U) = \frac{n}{r} \max(\| P_U e_i \|)$$

explain max n min
values of mu
mu = max(mu1,mu2)
do correction

- Weak coherence property, 2008.
- Strong coherence property, 2009. Result improved.

Mathematical results

Lemma: Information theoretic limit (C and Tao, 2009)

Coupon collector's effect, No method whatsoever can work if no of sampled enteries in the matrix is

$$m \lesssim \mu \times nr \times \log(n)$$

Candes and Recht, 2008

Recovering M exactly is possible with probability at least $1 - cn^{-}3$ when sampled enteries

$$m \gtrsim \mu \times n^{6/5} \times r\log(n) \qquad \text{sometimes } \boxed{(\tfrac{6}{5} \textit{ instead of } \tfrac{5}{4})}$$

Improving the result, Candes and Tao, 2009

Recovering M exactly is possible with probability at least $1 - n^{-}10$ when sampled enteries

$$m \gtrsim \mu \times n \times r\log^{a}(n) \qquad a \leq 6 \text{ (sometimes 2)}$$

If we have right $r(2n - r)$ numbers

If we chose the right numbers, we can hope to find only one matrix that will be the solution to the problem below.
Otherwise there could be multiple matrices that are consistant with those missing values and we can never know which one out of them all.

Rank minimisation problem

$$\text{minimize } \operatorname{rank}(X)$$
$$\text{subject to } X_{ij} = M_{ij} \quad \forall (i,j) \in \Omega_{known}$$

Conclusions

Not all directions are covered

- If there are more than one matrix that satisfies the above convex optimisation problem then it means you can't recover.
- It means the choice of number is not right.

How to minimise rank, NP hard

We were minimising the L_2 norm using gradient descent, which can be visualized as rank minimisation due to constraints. **(not sure though)**

$$min \parallel M' - X \parallel \quad \forall(i,j) \in \Omega_{known}$$

Nuclear norm

$$minimize \parallel X \parallel_*$$
$$subject\ to\ X_{ij} = M_{ij}\ \forall(i,j) \in \Omega_{known}$$
$$where, \parallel X \parallel_* = \sum_{i=1}^{r} \sigma_i^2$$

l_0 norm
l_1 norm

- This will be proxy for l_0 norm of singular values matrix.
- l_0 norm is number of non zero values, which is essentially rank of matrix.

Applications

- High dimentionality, but low rank structure.
  1. Netflix matrix
  2. Triangulation with sensor net.
  3. Quantum state tomography
  4. Machine learning
  5. Corrupted videos
- Robost PCA
  1. Recovering foreground from background
  2. Removing shadows from face images.
  3. Batch face alignment
- Transform invariant low rank textures.
- 3D reconstruction
- Latent semantic analysis