Mutual Coherence and The Welch Bound

Andersen Ang

Mathématique et de Recherche opérationnelle
Faculté polytechnique de Mons
UMONS
Mons, Belgium

email: manshun.ang@umons.ac.be
homepage: *angms.science*

August 6, 2017

Overview

Similarity measure between two non-zero vectors

Recall that the cosine distance c between two vector x and y has the following expression

$$c = \frac{x \cdot y}{\|x\|_2 \|y\|_2}$$

where $\cdot$ is the inner product and x, y are non-zero vectors in $\mathbb{R}^n$.

This expression comes from the Euclidean dot product

$$x \cdot y = \|x\|_2 \|y\|_2 \cos \theta$$

If x, y are normalized, their norms are 1, the expression of c becomes

$$c = x \cdot y$$

In other words, for unit vectors, "dot product" $=$ "correlation".

Mutual Coherence

Given a matrix $A \in \mathbb{R}^{m \times n}$ with normalized columns (i.e. $\|a_i\|_2 = 1$)

Mutual Coherence is defined to be the largest absolute value of correlation between the columns of A

$$M := \max_{i \neq j} |a_i \cdot a_j|$$

If the columns of A are not unit vectors, then mutual coherence is defined as

$$M := \max_{i \neq j} \frac{|a_i \cdot a_j|}{\|a_i\|_2 \|a_j\|_2}$$

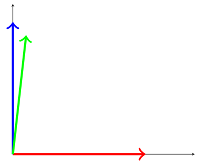
Note that $0 \leq M \leq 1$.
Quick proof: as norms and absolute value are non-negative, so $0 \leq M$.
For $M \leq 1$, consider Cauchy-Schwarz inequality for dot product :
$\|a_i \cdot a_j\| \leq \|a_i\|_2 \|a_j\|_2$, so $\dfrac{|a_i \cdot a_j|}{\|a_i\|_2 \|a_j\|_2} \leq \dfrac{\|a_i\|_2 \|a_j\|_2}{\|a_i\|_2 \|a_j\|_2} = 1$

What Mutual Coherence tells

Mutual-coherence M characterizes the dependence between columns of the matrix. Geometrically, mutual coherence tells how "spread" the two least spread vectors are: a small M means more "spread".

For example, consider the matrix

$$\begin{bmatrix} 1 & 0 & 0.1 \\ 0 & 1 & 0.9 \end{bmatrix}$$



Geometrically, we can see that $[0.1\ 0.9]^T$ and $[0\ 1]^T$ are cloest to each other $\iff$ smaller angle in between $\iff$ larger cosine value $\iff$ larger M

Computation of M

Given a matrix $X \in \mathbb{R}^{m \times n}$, recall that the Gram matrix $G = X^T X$ stores all the inner products between columns of X.

$$G = X^T X = \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \dots & x_1 \cdot x_n \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \dots & x_2 \cdot x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n \cdot x_1 & x_n \cdot x_2 & \dots & x_n \cdot x_n \end{bmatrix}$$

So M can be computed as the element with largest absolute value of G :

1. If X is not normalized, normalize it first.
2. Compute the gram matrix $G = X^T X$
3. Take absolute value on off-diagonal of G
4. Pick the largest value as M

The lower bound of Mutual Coherence - Welch inequality

In fact there is a lower bound of M for **any matrix** $A \in \mathbb{R}^{m \times n}$.

Theorem (Welch inequality). For a matrix $A \in \mathbb{R}^{m \times n}$ with $n > m$ (X is a short-fat matrix, having more columns than rows), the lower bound of M is :

$$M(m,n) \geq \sqrt{\frac{n-m}{m(n-1)}}$$

Note

- The bound is a function of m and n
- The bound is independent of the actually matrix A, so the inequality applies for all matrix A with size $m \times n$.

For example, when $m = 1$, $M \geq 1$, as $M \in [0\ 1]$ so $M = 1$. This is trivial as $m = 1$ means the column vectors are 1-dimensional points : either positive or negative , so $M = 1$.

Prerequisite for proving the Welch bound

We need the following to prove the Welch bound :

- Rank of gram matrix G of X equals to rank of X
  rank $X^T X =$ rank X.
- A matrix $X \in \mathbb{R}^{m \times n}$ has rank $\leq \min(m, n)$.
- A matrix $X \in \mathbb{R}^{m \times n}$ with rank r has r non-zero eigenvalues, all the other eigenvalues are zero.
- Cauchy Schwartz inequalities.
- Trace of a matrix is the sum of its diagonal elements.
- The mean of a set of non-negative numbers is smaller than their maximum. That is, for a group of numbers $a_1, a_2, ..., a_N \geq 0$, we have

$$\frac{1}{N} \sum_{i=1}^{N} a_i \leq \max a_i$$

The proof of lower bound of Mutual Coherence

Theorem (Welch inequality). For a matrix $X \in \mathbb{R}^{m \times n}$ with $n > m$ (X is a short-fat matrix having more columns than rows), the lower bound of M is :

$$M(m,n) \geq \sqrt{\frac{n-m}{m(n-1)}}$$

Proof. Consider a matrix $X \in \mathbb{R}^{m \times n}$ with normalized columns. The gram matrix is

$$G = X^T X = \begin{bmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \dots & x_1 \cdot x_n \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \dots & x_2 \cdot x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n \cdot x_1 & x_n \cdot x_2 & \dots & x_n \cdot x_n \end{bmatrix}$$

From the fact that rank $G =$ rank $X \leq \min(m,n) = m$, G has at most m eigenvalues, all the remaining eigenvalues are zeros.

The proof of lower bound of Mutual Coherence ... 2

Assume there are r eigenvalues of G, where $r \leq m < n$.
By the fact that the trace of a matrix is the sum of eigenvalues :

$$
\begin{aligned}
\text{Tr } G &= \sum_{i=1}^{r} \lambda_i \\
(\text{Tr } G)^2 &= \Big(\sum_{i=1}^{r} \lambda_i\Big)^2 \\
&= \Big(\sum_{i=1}^{r} \lambda_i \cdot 1\Big)^2 \\
\text{(Cauchy Schwartz)} &\leq \Big(\sum_{i=1}^{r} \lambda_i^2\Big)\Big(\sum_{i=1}^{r} 1^2\Big) \\
&= \Big(\sum_{i=1}^{r} \lambda_i^2\Big) r
\end{aligned}
$$

The proof of lower bound of Mutual Coherence ... 3

As $r \leq m$ and the remaining $n-r$ eigenvalues of G are all zeros, we have

$$(\text{Tr } G)^2 \quad \leq \quad m\Big(\sum_{i=1}^{n} \lambda_i^2\Big)$$

By the fact that for a matrix A we have the following inequlity :

$$\sum_{i=1}^{n} \lambda_i^2 \leq \sum_{ij} a_{ij}^2 = \|A\|_F^2$$

Hence for G

$$(\text{Tr } G)^2 \quad \leq \quad m\Big(\sum_{ij} (x_i \cdot x_j)^2\Big)$$

So

$$\sum_{ij} (x_i \cdot x_j)^2 \quad \geq \quad \frac{(\text{Tr}(G))^2}{m}$$

The proof of lower bound of Mutual Coherence ... 4

As x_i are unit vector, so the diagonal of G are all 1, hence

$$\operatorname{Tr} G = n \quad \text{and} \quad \sum_{ij}(x_i \cdot x_j)^2 = n + \sum_{i \neq j}(x_i \cdot x_j)^2$$

and therefore

$$\begin{aligned} \sum_{ij}(x_i \cdot x_j)^2 &\geq \frac{(\operatorname{Tr}(G))^2}{m} \\ n + \sum_{i \neq j}(x_i \cdot x_j)^2 &\geq \frac{n^2}{m} \\ \sum_{i \neq j}(x_i \cdot x_j)^2 &\geq \frac{n^2}{m} - n \\ &= \frac{n(n-m)}{m} \end{aligned}$$

The proof of lower bound of Mutual Coherence ... 5

A trick : the mean of a set of non-negative numbers is smaller than their maximum. That is, for a group of numbers $a_1, a_2, ..., a_N \geq 0$, we have

$$\frac{1}{N}\sum_{i=1}^{N} a_i \leq \max a_i$$

Recall that G is a n-by-n matrix and so is G^2. After removing the diagonal, the off-diagonal part $\sum_{i \neq j}(x_i \cdot x_j)^2$ has $n(n-1)$ non-negative $(x_i \cdot x_j)^2$ terms. Apply the trick we have

$$\frac{1}{n(n-1)}\sum_{i \neq j}(x_i \cdot x_j)^2 \leq \max(x_i \cdot x_j)^2$$

Put this into $\sum_{ij}(x_i \cdot x_j)^2 \geq \dfrac{n(n-m)}{m}$ and let $M = \max(x_i \cdot x_j)$ we get

$$\frac{1}{n(n-1)}\frac{n(n-m)}{m} \leq M^2 \implies M \geq \sqrt{\frac{n-m}{m(n-1)}}$$

□

Last page - summary

- Mutual coherence of a matrix A is $M := \max_{i \neq j} \frac{|a_i \cdot a_j|}{\|a_i\|_2 \|a_j\|_2}$
- $M =$ largest absolute value of coorelation between columns
- M charaterizes the degree of "spread" of the two least spread vectors
- Computation of M
- Welch bound : for matrix $A \in \mathbb{R}^{m \times n}$, $M \geq \sqrt{\frac{n-m}{m(n-1)}}$

End of document