Least squares method: linear regression

Ramani Duraiswami,
Dept. of Computer Science

Normal equations

$$2(A'Ac - A'y) = 0 \quad \text{or} \quad A'Ac = A'y$$

- The system is called the “Normal equations”
- Can solve least squares problems using these
- For $A$ size $m \times n$ and $c$ of size $n$ and $y$ of size $m$ what are the dimensions of the normal equations?
  - $n \times n$
- Have converted it to a regular system that we know how to solve
- Solve via LU decomposition
- Solution should be accurate if the matrix $A'A$ is well conditioned
More on Normal Equations

• Normal equations are only important theoretically
• Gives us a way to think about least squares.
• In practice least squares solved via a different process
  – QR decomposition
• Why?
  – Somewhat expensive as we have to form $A^t A$
  – involves matrix multiplication and then solution
  – More importantly it is poorly conditioned
  – $\text{cond}(A^t A) = (\text{cond}(A))^2$
• Would like a method whose errors are closer to the
  condition number of $A$

Look at the fitting matrix in more detail

• Suppose we want to solve via least squares
  \[ \mathbf{A} \mathbf{c} = \mathbf{y} \]
  – $\mathbf{A}$ is a $m \times n$ matrix with $m>n$
• One way to solve was via LU decomposition of normal
  equations
  – Poor condition numbers and so not recommended
  – Requires matrix-matrix multiplication which is expensive
• Instead
  – Look for methods that can directly operate on $\mathbf{A}$ to get the
    solution
  – Recall in LU we did a set of transformations to $\mathbf{A}$ and the r.h.s.
    to find $\mathbf{e}$
  – Today we will look at the QR algorithm
Vector Spaces

- A *linear combination* of vectors results in a new vector:

\[ \mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n \]

- If the only set of scalars such that

\[ \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n = \mathbf{0} \]

is \( \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0 \) then we say the vectors \( \{\mathbf{v}_i\} \) are *linearly independent*

- The *dimension* of a space is the greatest number of linearly independent vectors possible in a vector set

- For a vector space of dimension \( n \), any set of \( n \) linearly independent vectors form a *basis*

Linear Transformations: Matrices

- A *linear transformation*:
  - Maps one vector to another
  - Preserves linear combinations

- Turns out any linear transform can be represented by a *matrix*

- A \( M \times N \) matrix takes a vector with \( N \) elements to a vector with \( M \) elements
Key Ideas

- Column space of a matrix: the vector space formed by the collection of column vectors in a matrix.
- Every matrix vector product results in a vector formed by linear combination of vectors in the column space.
- A \( m \times n \) rectangular matrix \( A \) takes \( n \) vectors into \( m \) vectors.
- Let the least squares problem be \( Ac = f \).
- Let the solution which minimizes the residual be \( c_* \).
- Then \( c_* \) creates on matrix vector product a rhs \( f_* \) that is in the column space of \( A \).
- We want that \( c_* \) minimizes \( r = ||f - f_*|| \).

Null Space of A

- Not all \( m \) vectors will be reachable even if we supply arbitrary \( n \) vectors.
- \( \text{Range of } A \): the part of the space of \( m \) vectors that are reachable.
  \[
  \text{Range}(A) = \{ y \in R^m : y = Ax \text{ for some } x \in R^n \}
  \]
  - The range of \( A \) contains all those vectors that can be made up using the columns of \( A \).
  - \( \text{Rank}(A) \) is the dimension of the range of \( A \).
  - Null space of \( A \): those vectors \( x \), for which \( Ax \) is zero.
    \[
    \text{Null}(A) = \{ x \in R^n : Ax = 0 \}
    \]
  \[
  \text{Dim(Null}(A)) + \text{Rank}(A) = n
  \]
- Key idea: We want to minimize the error in the part that can be reached.
Null Space of $A^t$

- $A^t$ is a matrix that takes $m$ vectors into $n$ vectors
- Not all $n$ vectors may be reachable even if we supply arbitrary $m$ vectors
  - Range of $A^t$: the part of the space of $n$ vectors that are reachable
    \[ \text{Range}(A^t) = \{ y \in \mathbb{R}^n : y = A^t x \text{ for some } x \in \mathbb{R}^m \} \]
  - The range of $A^t$ contains all those vectors that can be made up using the rows of $A$
  - Rank($A^t$) is the dimension of the range of $A^t$
  - Null space of $A^t$: those vectors $x$, for which $A^t x$ is zero
    \[ \text{Null}(A^t) = \{ y \in \mathbb{R}^m : A^t y = 0 \} \]
    \[ \text{Dim}(\text{Null}(A^t)) + \text{Rank}(A^t) = m \]

QR decomposition

- Suppose we can write
  \[ A = Q' R' \]
  - $Q'$ is an orthonormal matrix of dimension $m \times m$
  - $R'$ is a $m \times n$ matrix that can be written as \[ [R] \]
    \[ [0] \]
  - $R$ is a triangular $n \times n$ matrix and $0$ is a matrix of zeroes of size $m-n \times n$
  - $Q'$ can also be partitioned as $[Q \; Q']$ with $Q$ containing $n$ orthonormal columns of size $m$ and $Q'_{m-n}$ orthonormal columns
- If $Ax = b$ then $(Q' \; R')x = b$ or $Q'(R' x) = b$ or $Q'y = b$
  - So if $b$ is in range($A$), it is also in range($Q'$)
  - Similarly if $Q'y = b$; then $b = Ax$ with $x = R^{-1}y$
  - Columns of $Q$ form an orthonormal basis for range($A$)
Orthogonal Matrices

- Orthogonal matrices are square matrices that have their columns orthonormal to each other
  - dot product of different column vectors is zero, while of the same column is one
    - Denoted $Q$
  - Most trivial orthogonal matrix is the identity matrix
    - $Q^t Q = \Lambda$
  - For an orthonormal matrix
    - $Q^t Q = I$
    - So $Q^{-1} = Q^t$

Generalization: a complex matrix is Hermitian iff $Q^{-1} = Q^H$ where superscript $^H$ denotes complex conjugate transpose

Orthogonal matrix facts

- Suppose $Q$ is an orthonormal matrix
- Then for any vector $r$ the Euclidean norm is preserved in an orthonormal transformation
- Proof
  $$\|Qr\|^2 = (Qr)^t (Qr) = r^t Q^t Q r = r^t (Q^t Q) r = r^t r = \|r\|^2$$
- If $Q$ is an orthonormal matrix
  - so is the extended matrix $Q_e$
  - $Q_e = \begin{bmatrix} I & 0 \\ 0 & Q \end{bmatrix}$
- Easy to show from definition that
  $$Q_e^t Q_e = I$$
Solving least squares with QR

• \( A = Q'R' \)

• Let \( r = y - Ac \) \( b = Q'^t y \)

• Goal of least squares find the \( c \) that minimizes squared error (residue)

• Partition \( b \) in to two pieces
  – \( b_1 \) of dimension \( n \)
  – \( b_2 \) of dimension \( m-n \)
  – \( ||r||^2 = ||y - Ac||^2 = ||y - Q' R' c||^2 \)
  – Length is not changed by multiplication with orthogonal matrix
  – So \( ||r||^2 = ||Q'^r||^2 = ||Q'^t [y - Q' R' c]||^2 \)
    \( = ||b_1 - R c||^2 + ||b_2 - 0c||^2 \)
So no matter what \( c \) is the second term remains unchanged
If we minimize \( ||r||^2 \) the best we can do is minimize first term

Solving LS via QR

• How do we minimize \( ||b_1 - R c||^2 \)
  – If \( R \) is full rank set solve \( Rc = b_1 \) then we have done the best we can
  – (if \( R \) is rank deficient solve in least squares sense)
  – Recall \( R \) is triangular so this equation can be easily solved

• Algorithm
  – Compute QR factorization of \( A = Q'R' \)
  – Form \( c_1 = Q^t b \)
  – Solve \( Rx = c_1 \)
  – If \( R \) is full rank and \( Q^{-1} \) is available then the norm of the residual is \( ||Q^{-1} b|| \). Else \( r = b - Ax \).