MATH 3795
Lecture 7. Linear Least Squares.

Dmitriy Leykekhman

Fall 2008

Goals

- Basic properties of linear least squares problems.
- Normal equation.

Linear Least Squares

- Given $A \in \mathbb{R}^{m \times n}$, we want to find $x \in \mathbb{R}^n$ such that $Ax \approx b$.
- If $m = n$ and A is invertible, then we can solve $Ax = b$.
- Otherwise, we may not have a solution of $Ax = b$ or we may have infinitely many of them.

Linear Least Squares

- Given $A \in \mathbb{R}^{m \times n}$, we want to find $x \in \mathbb{R}^n$ such that $Ax \approx b$.
- If $m = n$ and A is invertible, then we can solve $Ax = b$.
- Otherwise, we may not have a solution of $Ax = b$ or we may have infinitely many of them.
- We are interested in vectors x that minimize the norm of squares of the residual $Ax - b$, i.e., which solve

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

- The problems

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2, \quad \min_{x \in \mathbb{R}^n} \|Ax - b\|_2, \quad \frac{1}{2} \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

are equivalent in the sense that if x solves one of them it also solves the others.

Linear Least Squares

- Given $A \in \mathbb{R}^{m \times n}$, we want to find $x \in \mathbb{R}^n$ such that $Ax \approx b$.
- If $m = n$ and A is invertible, then we can solve $Ax = b$.
- Otherwise, we may not have a solution of $Ax = b$ or we may have infinitely many of them.
- We are interested in vectors x that minimize the norm of squares of the residual $Ax - b$, i.e., which solve

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

- The problems

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2, \quad \min_{x \in \mathbb{R}^n} \|Ax - b\|_2, \quad \frac{1}{2} \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

  are equivalent in the sense that if x solves one of them it also solves the others.
- Instead of finding x that minimizes the norm of squares of the residual $Ax - b$, we could also try to find x that minimizes the p-norm of the residual

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_p$$

  This can be done, but is more complicated and will not be covered.

Linear Least Squares

Example

Given m measurements

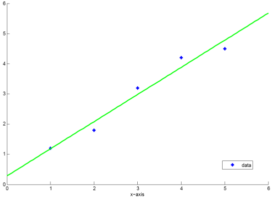
$$(x_i, y_i), \quad i = 1, \ldots, m,$$

find a linear function

$$y(x) = ax + b$$

that best fits these data, i.e.,

$$y_i \approx ax_i + b \quad i = 1, \ldots, m.$$



Linear Least Squares

We want two numbers a and b such that

$$\sum_{i=1}^{m}(ax_i + b - y_i)^2$$

is minimized.

Linear Least Squares

We want two numbers a and b such that

$$\sum_{i=1}^{m} (ax_i + b - y_i)^2$$

is minimized.

Write in matrix form. Let

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{pmatrix} \in \mathbb{R}^{m \times 2}, \quad b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m$$

then the ith residual

$$r_i = ax_i + b - y_i$$

is the ith component of $Az - b$, where $z = [a\ b]^T$. Thus we want to minimize $\|r\|_2^2$ which leads to

$$\min_{z \in \mathbb{R}^2} \|Az - b\|_2^2$$

Linear Least Squares

Example

more generally, given m measurements

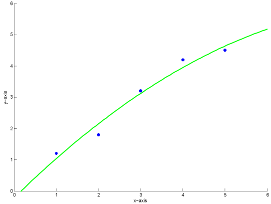
$$(x_i, y_i), \quad i = 1, \ldots, m,$$

find a polynomial function

$$y(x) = a_n x^n + \cdots \; a_1 x + a_0$$

that best fits these data, i.e.,

$$y_i \approx a_n x_i^n + \cdots \; a_1 x_i + a_0 \quad i = 1, \ldots, m.$$



Linear Least Squares

We want two numbers a and b such that

$$\sum_{i=1}^{m}(a_n x_i^n + \cdots \; a_1 x_i + a_0 - y_i)^2$$

is minimized.

Linear Least Squares

We want two numbers a and b such that

$$\sum_{i=1}^{m} (a_n x_i^n + \cdots \ a_1 x_i + a_0 - y_i)^2$$

is minimized.

Write in matrix form. Let

$$A = \begin{pmatrix} x_1^n & \dots & x_1 & 1 \\ x_2^n & \dots & x_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_m^n & \dots & x_m & 1 \end{pmatrix} \in \mathbb{R}^{m \times n}, \quad b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m$$

then the ith residual

$$r_i = a_n x_i^n + \cdots \ a_1 x_i + a_0 - y_i$$

is the ith component of $Az - b$, where $z = [a_n, \ \dots, \ a_1, \ a_0]^T$. Thus we want to minimize $\|r\|_2^2$ which leads again to

$$\min_{z \in \mathbb{R}^n} \|Az - b\|_2^2$$

Linear Least Squares

Example

Find a best fit circle through points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$.
Equation for the circle around (c_1, c_2) with radius r is

$$(x - c_1)^2 + (y - c_2)^2 = r^2.$$

Linear Least Squares

Example

Find a best fit circle through points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$.
Equation for the circle around (c_1, c_2) with radius r is

$$(x - c_1)^2 + (y - c_2)^2 = r^2.$$

Rewrite the equation for the circle in the form

$$2xc_1 + 2yc_2 + (r^2 - c_1^2 - c_2^2) = x^2 + y^2.$$

Linear Least Squares

Example

Find a best fit circle through points $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$. Equation for the circle around (c_1, c_2) with radius r is

$$(x - c_1)^2 + (y - c_2)^2 = r^2.$$

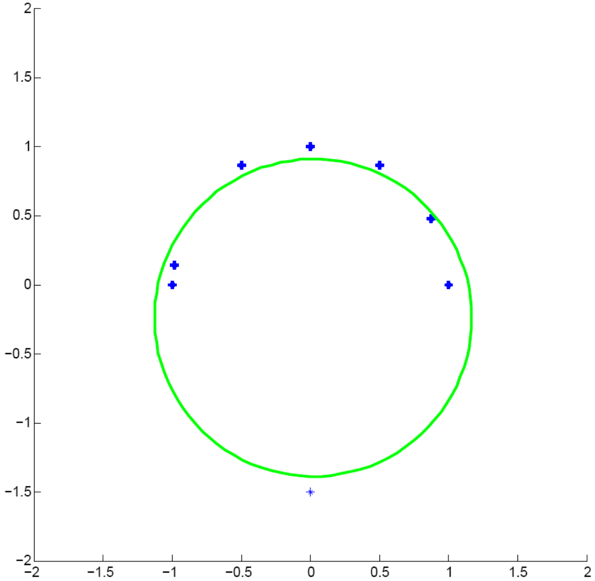
Rewrite the equation for the circle in the form

$$2xc_1 + 2yc_2 + (r^2 - c_1^2 - c_2^2) = x^2 + y^2.$$

Set $c_3 = r^2 - c_1^2 - c_2^2$, then we can compute the center (c_1, c_2) and the radius $r = \sqrt{c3 + c_1^2 + c_2^2}$ of the circle that best fits the data points by solving the least squares problem

$$\min_{[c_1,c_2,c_3]^T \in \mathbb{R}^3} \left\| \begin{pmatrix} 2x_1 & 2y_1 & 1 \\ 2x_2 & 2y_2 & 1 \\ \vdots & \vdots & \vdots \\ 2x_m & 2y_m & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} - \begin{pmatrix} x_1^2 + y_1^2 \\ x_2^2 + y_2^2 \\ \vdots \\ x_m^2 + y_m^2 \end{pmatrix} \right\|_2^2$$

Linear Least Squares



Linear Least Squares

- Suppose x_* satisfies

$$\|Ax_* - b\|_2^2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 \quad \text{(LLS)}$$

- For any vector $z \in \mathbb{R}^n$

$$\begin{aligned}
&\|Ax_* - b\|_2^2 \leq \|A(x_* + z) - b\|_2^2 \\
&= (A(x_* + z) - b)^T (A(x_* + z) - b) \\
&= x_*^T A^T A x_* - 2x_*^T A^T b + b^T b + 2z^T A^T A x_* - 2z^T A^T b + z^T A^T z \\
&= \|Ax_* - b\|_2^2 + 2z^T (A^T A x_* - A^T b) + \|Az\|_2^2.
\end{aligned}$$

Linear Least Squares

- Suppose x_* satisfies

$$\|Ax_* - b\|_2^2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 \quad \text{(LLS)}$$

- For any vector $z \in \mathbb{R}^n$

$$\begin{aligned}
&\|Ax_* - b\|_2^2 \leq \|A(x_* + z) - b\|_2^2 \\
&= (A(x_* + z) - b)^T (A(x_* + z) - b) \\
&= x_*^T A^T A x_* - 2x_*^T A^T b + b^T b + 2z^T A^T A x_* - 2z^T A^T b + z^T A^T z \\
&= \|Ax_* - b\|_2^2 + 2z^T (A^T A x_* - A^T b) + \|Az\|_2^2.
\end{aligned}$$

- Of course $\|Az\|_2^2 \geq 0$, but

$$2z^T (A^T A x_* - A^T b)$$

could be negative for some z if $A^T A x_* - A^T b \neq 0$.

Linear Least Squares

- In fact setting
$$z = -\alpha(A^T A x_* - A^T b)$$
for some $\alpha \in \mathbb{R}$
- For such $z \in \mathbb{R}^n$ we get
$$\begin{aligned} &2z^T(A^T A x_* - A^T b) + \|Az\|_2^2 \\ &= -2\alpha\|A^T A x_* - A^T b\|_2^2 + \alpha^2\|A(A^T A x_* - A^T b)\|_2^2 < 0 \end{aligned}$$
for
$$0 < \alpha < \frac{\|A^T A x_* - A^T b\|_2^2}{\|A(A^T A x_* - A^T b)\|_2^2}.$$

Linear Least Squares

- In fact setting

$$z = -\alpha(A^T A x_* - A^T b)$$

for some $\alpha \in \mathbb{R}$

- For such $z \in \mathbb{R}^n$ we get

$$\begin{aligned} &2z^T(A^T A x_* - A^T b) + \|Az\|_2^2 \\ &= -2\alpha\|A^T A x_* - A^T b\|_2^2 + \alpha^2\|A(A^T A x_* - A^T b)\|_2^2 < 0 \end{aligned}$$

for

$$0 < \alpha < \frac{\|A^T A x_* - A^T b\|_2^2}{\|A(A^T A x_* - A^T b)\|_2^2}.$$

- Thus, if x_* solves (LLS) then x_* must satisfy

$$A^T A x_* - A^T b = 0 \quad \textbf{normal equation}.$$

Linear Least Squares

On the other hand if x_* satisfies

$$A^T A x_* - A^T b = 0,$$

then for any x

$$\begin{aligned}
\|Ax - b\|_2^2 &= \|Ax_* + A(x - x_*) - b\|_2^2 \\
&= \|Ax_* - b\|_2^2 + 2(x - x_*)^T (A^T A x_* - A^T b) + \|A(x - x_*)\|_2^2 \\
&= \|Ax_* - b\|_2^2 + \|A(x - x_*)\|_2^2 \\
&\geq \|Ax_* - b\|_2^2
\end{aligned}$$

i.e. x_* solves (LLS).

Linear Least Squares. Normal Equation.

Theorem

The linear least square problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 \quad \textit{(LLS)}$$

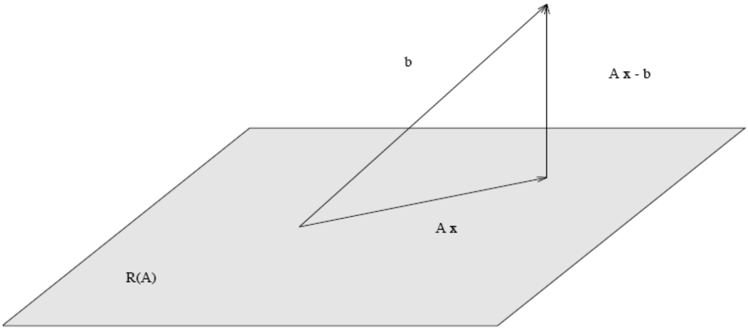
always has a solution. A vector x_* solves (LLS) iff x_* solves the normal equation

$$A^T A x = A^T b.$$

Note: If the matrix $A \in \mathbb{R}^{m \times n}$, $m \geq n$, has rank n, then $A^T A$ is symmetric positive definite and satisfies

$$v^T A^T A v = \|Av\|_2^2 > 0, \quad \forall v \in \mathbb{R}^n, \; v \neq 0.$$

Linear Least Squares



Linear Least Squares. Normal Equation.

If $A \in \mathbb{R}^{m \times n}$, $m \geq n$, has full rank n, then we can use the Cholesky-decomposition to solve the normal equation (and, hence, the linear least squares problem) as follows

1. Compute $A^T A$ and $A^T b$.
2. Compute the Cholesky-decomposition $A^T A = R^T R$.
3. Solve $R^T y = A^T b$ (forward solve),
   solve $Rx = y$ (backward solve) .

The computation of $A^T A$ and $A^T b$ requires roughly mn^2 and $2mn$ flops. Roughly $\frac{1}{3}n^3$ flops are required to compute the Cholesky-decomposition. The solution of $R^T y = A^T b$ and of $Rx = y$ requires approximately $2n^2$ flops.

Linear Least Squares. Normal Equation.

Computing the normal equations requires us to calculate terms of the form $\sum_{k=1}^{m} a_{ki}a_{kj}$. The computed matrix A^TA may not be positive definite, because of floating point arithmetic.

```
t = 10.^(0:-1:-10)';
A = [ ones(size(t)) t t.^2 t.^3 t.^4 t.^5];
B = A'*A;
[R,iflag] = chol( B );
if( iflag ~= 0 )
disp([' Cholesky decomposition returned with iflag = ', ...
int2str(iflag)])
end
```

In exact arithmetic $B = A^TA$ is symmetric positive definite, but the Cholesky-Decomposition detects that $a_{jj} - \sum_{k=1}^{j-1} r_{jk}^2 < 0$ in step $j = 6$.

```
>> Cholesky decomposition returned with iflag = 6
```

The use of the Cholesky decomposition is problematic if the condition number of A^TA is large. In the example, $\kappa_2(A^TA) \approx 4.7 * 10^{16}$.