Sistemas lineales sobredeterminados

Transparencias basadas en el libro *Computing: An Introductory Survey* by Michael T. Heath, copyright © 2018 by the Society for Industrial and Applied Mathematics. http://www.siam.org/books/cl80

Mínimos cuadrados lineales

Método de mínimos cuadrados

- Los errores de medición y otras fuentes de variación aleatoria son inevitables en las ciencias experimentales.

- Dicha variabilidad se puede suavizar promediando muchos casos, por ejemplo, tomando más medidas de las estrictamente necesarias para determinar los parámetros del sistema.
- El sistema resultante está sobredeterminado, por lo que normalmente no hay una solución exacta
- Los datos de dimensiones superiores se proyectan en un espacio de dimensiones inferiores para suprimir el ruido o los detalles irrelevantes.
- Tal proyección la calculamos por el método de mínimos cuadrados

Mínimos cuadrados lineales

- En problemas lineales obtenemos un sistema lineal sobredeterminado $\boldsymbol{Ax} = \boldsymbol{b}$, con $\boldsymbol{A}$ **una matriz m × n**, $m > n$
- El sistema es más correcto poner $\boldsymbol{Ax} \cong \boldsymbol{b}$, dado que la igualdad no se suele satisfacer cuando $m > n$
- La solución de mínimos cuadrados $\boldsymbol{x}$ minimiza la norma Euclidea del vector $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{Ax}$,

$$\min_{\boldsymbol{x}} \|\boldsymbol{r}\|_2^2 = \min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{Ax}\|_2^2$$

Ajuste de datos

- Dados m datos (t_i, y_i), debemos encontrar el n-vector x de parámetros que da el "mejor ajuste" a la función del modelo$f(t, \boldsymbol{x})$,

$$\min_{\boldsymbol{x}} \sum_{i=1}^{m} (y_i - f(t_i, \boldsymbol{x}))^2$$

- El problema es *lineal* si la función f es una combinación lineal de las componentes de **x**

$$f(t, \boldsymbol{x}) = x_1\phi_1(t) + x_2\phi_2(t) + \cdots + x_n\phi_n(t)$$

  donde las funciones ϕ_j solo dependen de t
- $\cong$
  El problema lineal se puede escribir matricialmente como $\boldsymbol{Ax}$ $=\boldsymbol{b}$, con$a_{ij} = \phi_j(t_i)$ y $b_i = y_i$

Ajuste de datos

- Ajuste polinomial

$$f(t, \boldsymbol{x}) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

  es lineal, ya que el polinomio es lineal en sus coeficientes, aunque no lineal en la variable independiente t

- Suma de ajustes exponenciales

$$f(t, \boldsymbol{x}) = x_1 e^{x_2 t} + \cdots + x_{n-1} e^{x_n t}$$

  es un ejemplo de problema no lineal

- Por ahora, consideraremos solo problemas de mínimos cuadrados lineales

Example: Data Fitting

- Ajustar un polinomio cuadrático a cinco puntos de datos da un problema de mínimos cuadrados lineales

$$\boldsymbol{Ax} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cong \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \boldsymbol{b}$$

- La matriz cuyas columnas (o filas) son potencias sucesivas de la variable independiente se llamamatriz de V*andermonde*

Example,

- Para los datos

t	-1.0	-0.5	0.0	0.5	1.0
y	1.0	0.5	0.0	0.5	2.0

El sistema lineal sobredeterminado de tamaño 5x3 es

$$\boldsymbol{Ax} = \begin{bmatrix} 1 & -1.0 & 1.0 \\ 1 & -0.5 & 0.25 \\ 1 & 0.0 & 0.0 \\ 1 & 0.5 & 0.25 \\ 1 & 1.0 & 1.0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cong \begin{bmatrix} 1.0 \\ 0.5 \\ 0.0 \\ 0.5 \\ 2.0 \end{bmatrix} = \boldsymbol{b}$$

- La solución, que veremos más adelante cómo calcularla es

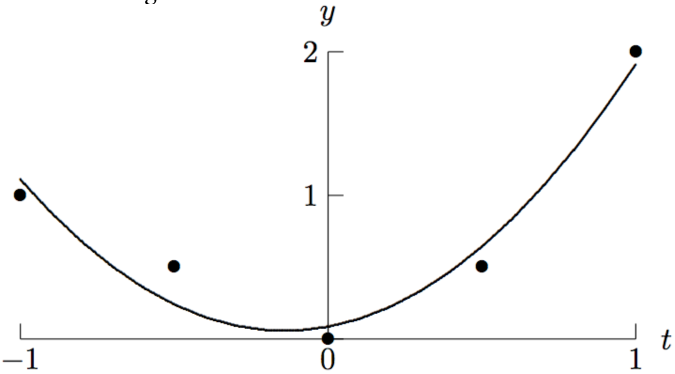
$$\boldsymbol{x} = \begin{bmatrix} 0.086 & 0.40 & 1.4 \end{bmatrix}^T$$

entonces el polinomio de aproximación es

$$p(t) = 0.086 + 0.4t + 1.4t^2$$

Example,

- La curva resultante y los puntos correspondientes a los datos originales se muestran en el gráfico



Existencia y unicidad

Existencia y unicidad

- El problema de mínimos cuadrados lineales $\boldsymbol{Ax} \cong \boldsymbol{b}$ siempre tiene solución
- La solución es única si, y sólo si, las columnas de $\boldsymbol{A}$ son linealmente independientes, i.e., $\text{rank}(\boldsymbol{A}) = n$, con $\boldsymbol{A}$ ide tamaño $m \times n$
- Si $\text{rank}(\boldsymbol{A}) < n$, entonces $\boldsymbol{A}$ es de rango deficiente, y la solución del ajuste por mínimos cuadrados no es único
- A partir de ahora, asumimos que $\boldsymbol{A}$ tiene rango completo por columnas $\text{rank } n$

Normal Equations

- Para minimizar la norma euclidia al cuadrado del vector residual

$$\begin{aligned} \|\boldsymbol{r}\|_2^2 \quad &= \quad \boldsymbol{r}^T\boldsymbol{r} = (\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}) \\ &= \boldsymbol{b}^T\boldsymbol{b} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} \end{aligned}$$

Calculamos la derivada respecto $\boldsymbol{x}$ y la igualamos a $\mathbf{0}$,

$2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{A}^T\boldsymbol{b} =$ ecuaciones normales

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$$

Ortogonalidad

- Los vectores $\boldsymbol{v}_1$ y $\boldsymbol{v}_2$ son *ortogonales* si el producto escalar , $\boldsymbol{v}^T$ $\boldsymbol{v}_2 = 0$
  $\boldsymbol{v}_1$

- El espacio generados por las columnas de la matriz **A** m × n

- ,
  $\text{span}(\boldsymbol{A}) = \{\boldsymbol{Ax} : \boldsymbol{x} \in \mathbb{R}^n\}$,tiene dimensión como mucho n

- Si $m > n$, $\boldsymbol{b}$ generalmente no están en $\text{span}(\boldsymbol{A})$, por lo que no existe solución exacta de $\boldsymbol{Ax} = \boldsymbol{b}$

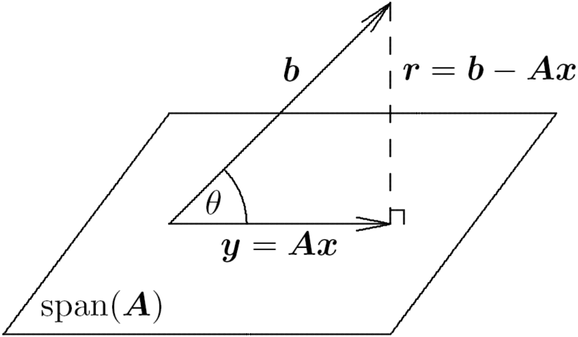
El vector $\boldsymbol{y} = \boldsymbol{Ax}$ está en $\text{span}(\boldsymbol{A})$ próximo a $\boldsymbol{b}$ respecto la norm-2 cuando el vector residual $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{Ax}$ is *ortogonal* a $\text{span}(\boldsymbol{A})$,

$$\boldsymbol{0} = \boldsymbol{A}^T\boldsymbol{r} = \boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{Ax})$$
$$\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{A}^T\boldsymbol{b}$$

De nuevo obtenemos las ecuaciones normales

Ortogonalidad

- La relación geométrica entre $\boldsymbol{b}$, $\boldsymbol{r}$, y span($\boldsymbol{A}$) se muestra en el siguiente gráfico



Matrices ortogonales

- La matriz $\boldsymbol{P}$ es ortogonal si es idempotente ($\boldsymbol{P}^2 = \boldsymbol{P}$) y simétrica($\boldsymbol{P}^T = \boldsymbol{P}$)

- La matriz ortogonal y el complemento ortogonal $\mathrm{span}(\boldsymbol{P})^\perp$ viene dado por $\boldsymbol{P}_\perp = \boldsymbol{I} - \boldsymbol{P}$

- Cualquier $\boldsymbol{v}$,

$$\boldsymbol{v} = (\boldsymbol{P} + (\boldsymbol{I} - \boldsymbol{P}))\,\boldsymbol{v} = \boldsymbol{P}\boldsymbol{v} + \boldsymbol{P}_\perp \boldsymbol{v}$$

- El problema de mínimos cuadrados $\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b}$, si $\mathrm{rank}(\boldsymbol{A}) = n$, entonces

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$$

es la proyección ortogonal sobre el subespacio $\mathrm{span}(\boldsymbol{A})$, y

$$\boldsymbol{b} = \boldsymbol{P}\boldsymbol{b} + \boldsymbol{P}_\perp \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} + (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{y} + \boldsymbol{r}$$

Pseudoinversa y el Número de Condición

- La matriz rectangular $\mathbf{A}$ $m \times n$ no tiene inversa en el sentido clásico

- lSi $\operatorname{rank}(\boldsymbol{A}) = n$, la pseudoinversa se define por

$$\boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$$

  y el número de condición

$$\operatorname{cond}(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{A}^+\|_2$$

- By convention, $\operatorname{cond}(\boldsymbol{A}) = \infty$ if $\operatorname{rank}(\boldsymbol{A}) < n$

- Al igual que el número condición de una matriz cuadrada mide lo próxima que está a ser singular, en las matrices rectangulares mide lo próxima que está de tener rango deficienteJ

- La solución de mínimos cudrado $\boldsymbol{Ax} \cong \boldsymbol{b}$ viene dada por $\boldsymbol{x} = \boldsymbol{A}^+\,\boldsymbol{b}$

Sensibilidad y condicionamiento

- La sensibilidad de la solución de mínimos cuadrado $\boldsymbol{Ax} \cong \boldsymbol{b}$ depende de $\boldsymbol{b}$ así como de $\boldsymbol{A}$

- Se define el ángulo θ entre $\boldsymbol{b}$ y $\boldsymbol{y} = \boldsymbol{Ax}$ por

$$\cos(\theta) = \frac{\|\boldsymbol{y}\|_2}{\|\boldsymbol{b}\|_2} = \frac{\|\boldsymbol{Ax}\|_2}{\|\boldsymbol{b}\|_2}$$

- Si limitamos la perturbación en la solución $\mathbf{x}$,$\Delta\boldsymbol{x}$ pro la perturbación $\Delta\boldsymbol{b}$ en $\boldsymbol{b}$, viene dada por

$$\frac{\|\Delta\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leq \operatorname{cond}(\boldsymbol{A}) \frac{1}{\cos(\theta)} \frac{\|\Delta\boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2}$$

Sensibilidad y condicionamiento

- Análogamente, respecto a la perturbación $\boldsymbol{E}$ en la matriz $\boldsymbol{A}$,

$$\frac{\|\Delta \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \lessapprox \left([\mathrm{cond}(\boldsymbol{A})]^2 \tan(\theta) + \mathrm{cond}(\boldsymbol{A})\right) \frac{\|\boldsymbol{E}\|_2}{\|\boldsymbol{A}\|_2}$$

- El número de condición de la solución de mínimos cuadrados del orden del cond(A) si el residual es pequeño, pero puede elevarse al cuadrado o arbitrariamente peor para un residual grande

Resolviendo problemas de mínimos cuadrados

Método de las ecuaciones normales

- Si la matriz $\boldsymbol{A}$ de tamaño mxn tiene rango n, entonces la matriz simétrica $\boldsymbol{A}^T\boldsymbol{A}$ de tamaño nxn es definida positiva, por lo tanto, la factorización de Cholesky

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$$

puede utilizarse pararesolver el sistema de ecuaciones normales

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$$

que tiene la misma solución que el sistema de mínimos cuadrados $\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b}$

- Las ecuaciones normales llevan involucradas las siguientes transformaciones

$$\text{rectangular} \longrightarrow \text{cuadrada} \longrightarrow \text{triangular}$$

que preservan la solución de mínimos cuadrados en principio, pero pueden no ser satisfactorias en aritmética de precisión finita.

Example: Método de las ecuaciones normales

- Para el ejemplo de ajuste de datos polinómicos dado anteriormente, el método de ecuaciones normales da

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1.0 & -0.5 & 0.0 & 0.5 & 1.0 \\ 1.0 & 0.25 & 0.0 & 0.25 & 1.0 \end{bmatrix} \begin{bmatrix} 1 & -1.0 & 1.0 \\ 1 & -0.5 & 0.25 \\ 1 & 0.0 & 0.0 \\ 1 & 0.5 & 0.25 \\ 1 & 1.0 & 1.0 \end{bmatrix}$$

$$= \begin{bmatrix} 5.0 & 0.0 & 2.5 \\ 0.0 & 2.5 & 0.0 \\ 2.5 & 0.0 & 2.125 \end{bmatrix},$$

$$\boldsymbol{A}^T\boldsymbol{b} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1.0 & -0.5 & 0.0 & 0.5 & 1.0 \\ 1.0 & 0.25 & 0.0 & 0.25 & 1.0 \end{bmatrix} \begin{bmatrix} 1.0 \\ 0.5 \\ 0.0 \\ 0.5 \\ 2.0 \end{bmatrix} = \begin{bmatrix} 4.0 \\ 1.0 \\ 3.25 \end{bmatrix}$$

Example,

- La factorización de Cholesky de la matriz simétrica definida positiva $\boldsymbol{A}^T\boldsymbol{A}$ da

$$
\begin{aligned}
\boldsymbol{A}^T\boldsymbol{A} &= \begin{bmatrix} 5.0 & 0.0 & 2.5 \\ 0.0 & 2.5 & 0.0 \\ 2.5 & 0.0 & 2.125 \end{bmatrix} \\
&= \begin{bmatrix} 2.236 & 0 & 0 \\ 0 & 1.581 & 0 \\ 1.118 & 0 & 0.935 \end{bmatrix} \begin{bmatrix} 2.236 & 0 & 1.118 \\ 0 & 1.581 & 0 \\ 0 & 0 & 0.935 \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^T
\end{aligned}
$$

- Resolviendo el sistema triangular $\boldsymbol{L}\boldsymbol{z} = \boldsymbol{A}^T\boldsymbol{b}$ por sustitución hacia atras da $\boldsymbol{z} = \quad 1.789 \quad 0.632 \quad 1.336 \quad]^T$

- Resolviendo el sistema triangular superior $\mathbf{L}^T\boldsymbol{x} = \boldsymbol{z}$ da $\boldsymbol{x} = [0.086 \quad 0.400 \quad 1.429$

Deficiencias de las ecuaciones normales

- Cierta información puede perderse al calucular $\boldsymbol{A}^T\boldsymbol{A}$ y $\boldsymbol{A}^T\boldsymbol{b}$

- Por ejemplo, considera

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ \epsilon & 0 \\ 0 & \epsilon \end{bmatrix}$$

  donde ϵ es unnúmero positivo más pequeño que $\sqrt{\epsilon_{\text{mach}}}$

- Entonces en aritmética flotante

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} 1+\epsilon^2 & 1 \\ 1 & 1+\epsilon^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

  que es singular

- La sensibilidad de la solución también empeora, ya que

$$\text{cond}(\boldsymbol{A}^T\boldsymbol{A}) = [\text{cond}(\boldsymbol{A})]^2$$