Lecture 9 – Linear Least Squares

Dr. Qingquan Sun

School of Computer Science and Engineering
Cal State University San Bernardino

Residual

- *Residual vector* of approximate solution $\hat{x}$ to linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is defined by

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}$$

- In theory, if $\boldsymbol{A}$ is nonsingular, then $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\| = 0$ if, and only if, $\|\boldsymbol{r}\| = 0$, but they are not necessarily small simultaneously

- Since

$$\frac{\|\Delta \boldsymbol{x}\|}{\|\hat{\boldsymbol{x}}\|} \leq \operatorname{cond}(\boldsymbol{A}) \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{A}\| \cdot \|\hat{\boldsymbol{x}}\|}$$

  small relative residual implies small relative error in approximate solution *only if* $\boldsymbol{A}$ is well-conditioned

Least Square Method

- Measurement errors are inevitable in observational and experimental sciences
- Errors can be smoothed out by averaging over many cases, i.e., taking more measurements than are strictly necessary to determine parameters of system

Least Square Method

- Measurement errors are inevitable in observational and experimental sciences
- Errors can be smoothed out by averaging over many cases, i.e., taking more measurements than are strictly necessary to determine parameters of system
- Resulting system is *overdetermined*, so usually there is no exact solution

Least Square Method

- Measurement errors are inevitable in observational and experimental sciences
- Errors can be smoothed out by averaging over many cases, i.e., taking more measurements than are strictly necessary to determine parameters of system
- Resulting system is *overdetermined*, so usually there is no exact solution
- In effect, higher dimensional data are projected into lower dimensional space to suppress irrelevant detail
- Such projection is most conveniently accomplished by method of *least squares*

Least Square Method

- For linear problems, we obtain *overdetermined* linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, with $m \times n$ matrix $\boldsymbol{A}$, $m > n$
- System is better written $\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b}$, since equality is usually not exactly satisfiable when $m > n$

Least Square Method

- For linear problems, we obtain *overdetermined* linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, with $m \times n$ matrix $\boldsymbol{A}$, $m > n$
- System is better written $\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b}$, since equality is usually not exactly satisfiable when $m > n$
- Least squares solution $\boldsymbol{x}$ minimizes squared Euclidean norm of residual vector $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$,

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

Data Fitting

- Given m data points (t_i, y_i), find n-vector $\boldsymbol{x}$ of parameters that gives "best fit" to model function $f(t, \boldsymbol{x})$,

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

- Problem is *linear* if function f is linear in components of $\boldsymbol{x}$,

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

  where functions ϕ_j depend only on t

- Problem can be written in matrix form as $\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b}$, with $a_{ij} = \phi_j(t_i)$ and $b_i = y_i$

Data Fitting

- Polynomial fitting

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

is linear, since polynomial linear in coefficients, though nonlinear in independent variable t

- Fitting sum of exponentials

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

is example of nonlinear problem

- For now, we will consider only linear least squares problems

Data Fitting Example

- Fitting quadratic polynomial to five data points gives linear least squares problem

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

Data Fitting Example

- For data

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

overdetermined 5×3 linear system is

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

- Solution, which we will see later how to compute, is

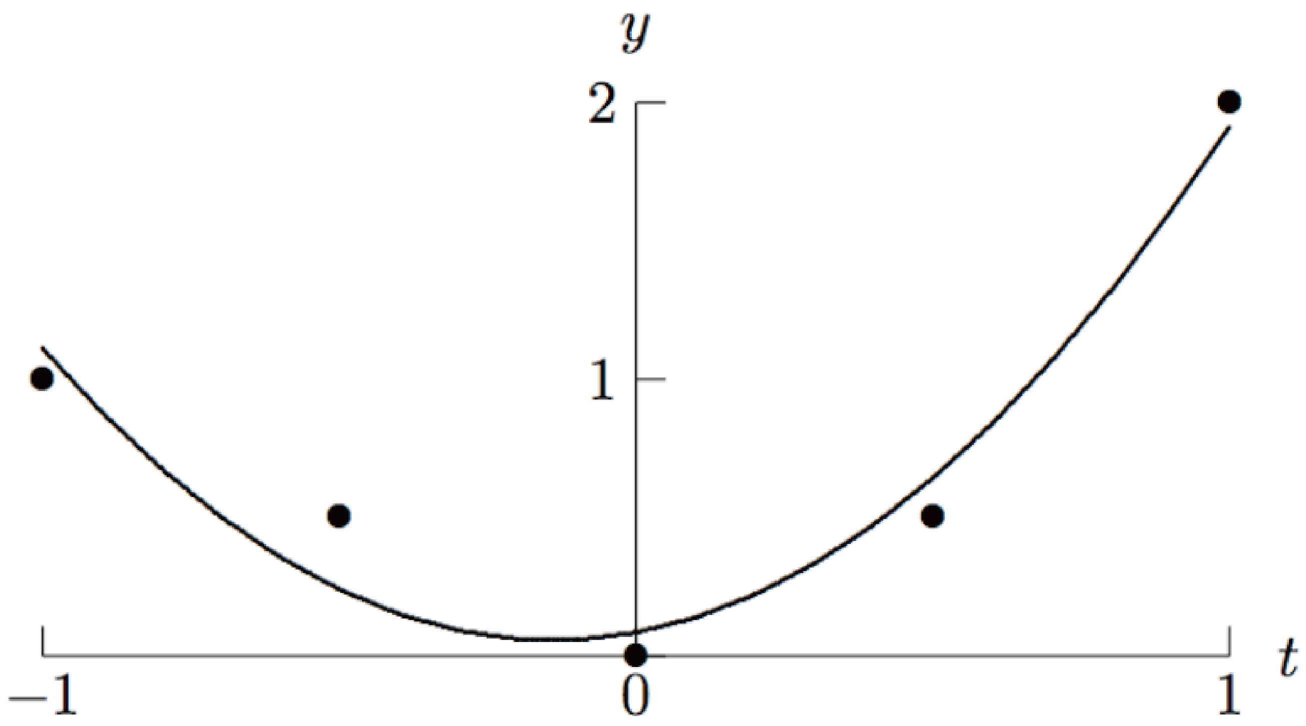
$$\boldsymbol{x} = \begin{bmatrix} 0.086 & 0.40 & 1.4 \end{bmatrix}^T$$

so approximating polynomial is

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

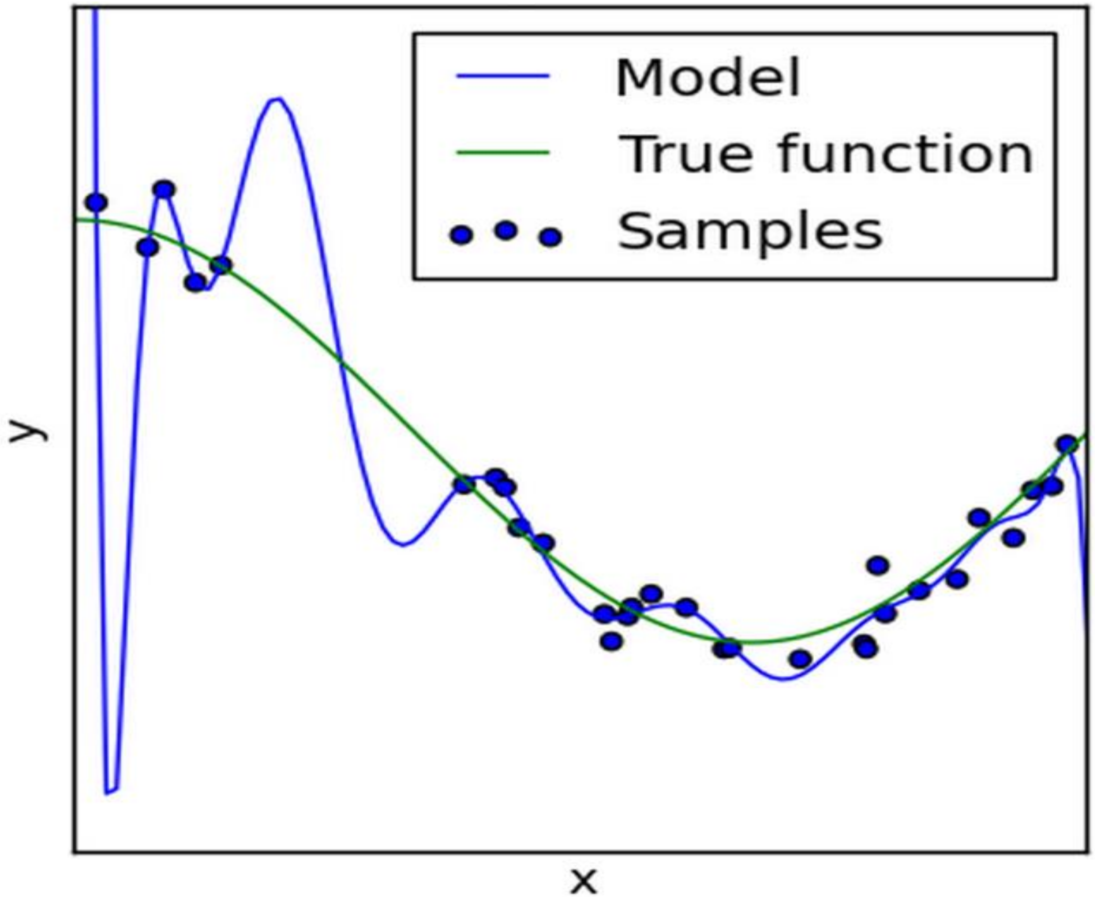
Data Fitting Example

- Resulting curve and original data points are shown in graph



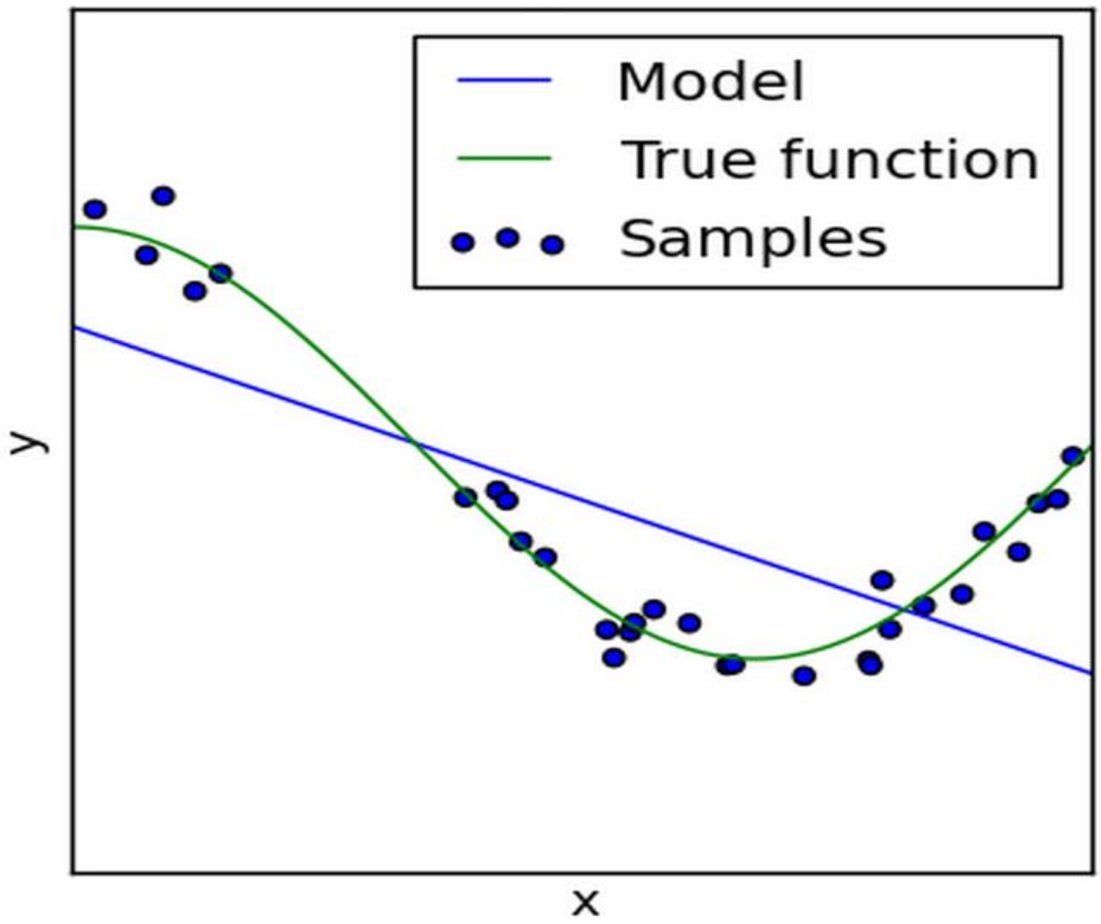
Over-fitting

- A model describes noise or error instead of the underlying relationship

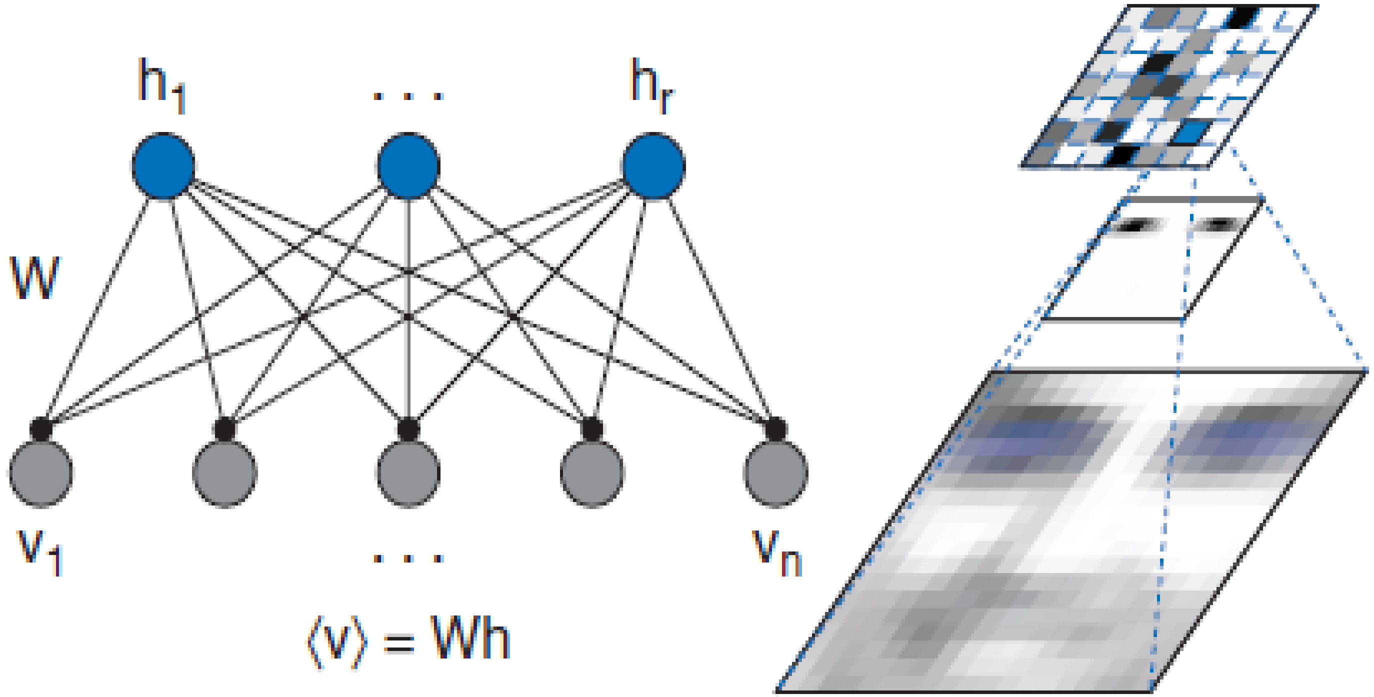


Under-fitting

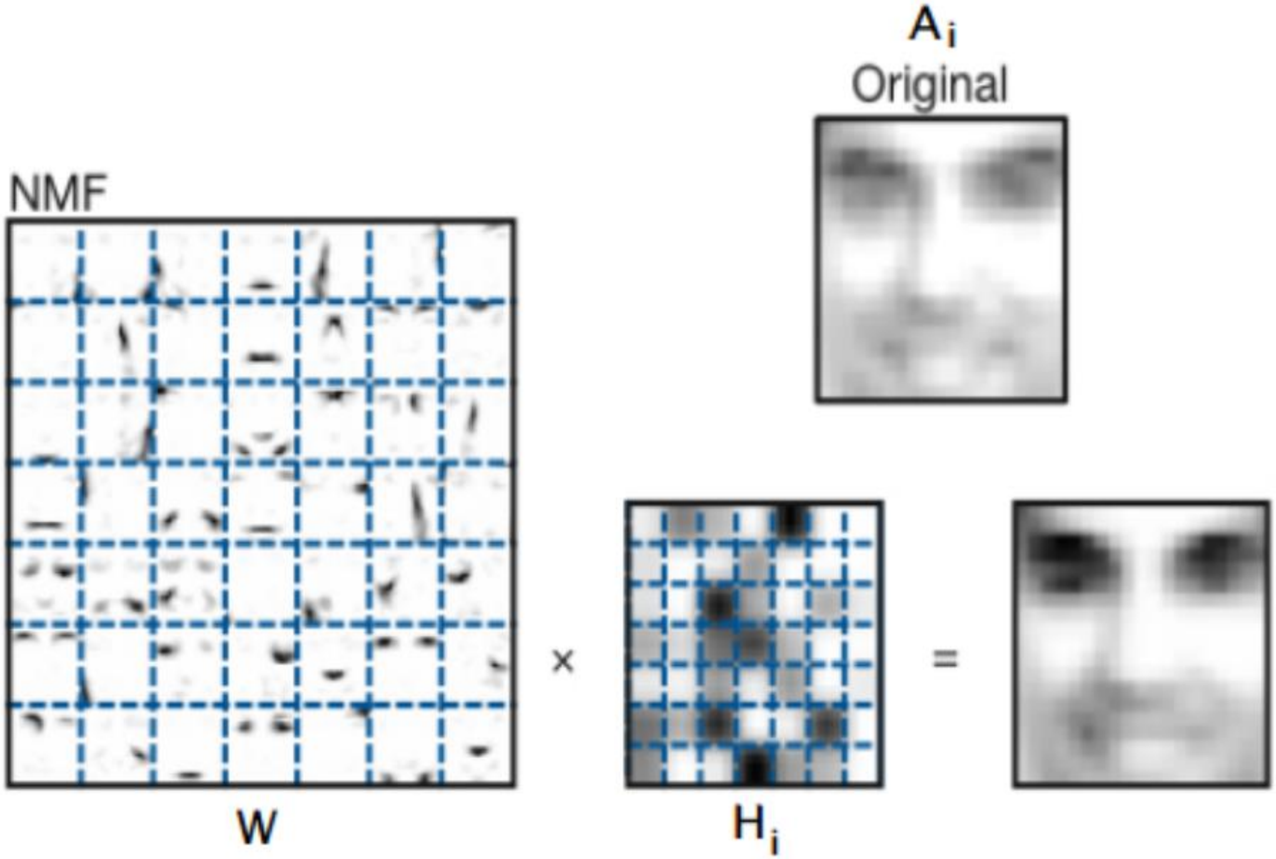
- A model fails to describe the underlying relationship



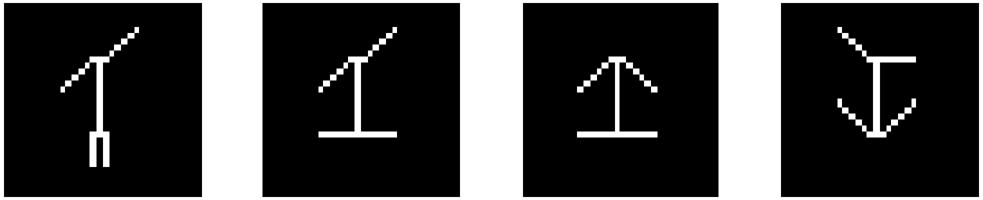
Nonnegative Matrix Factorization



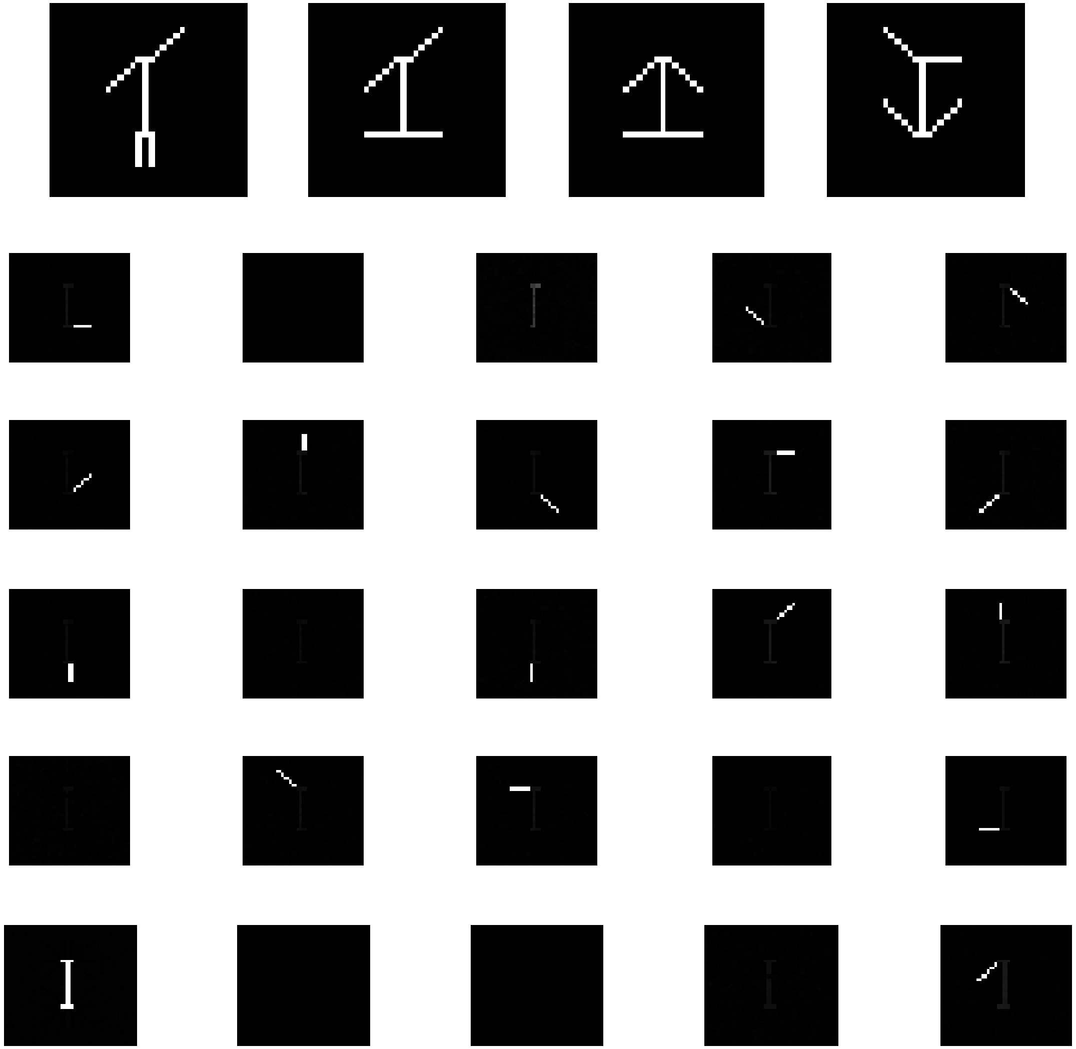
Nonnegative Matrix Factorization



Nonnegative Matrix Factorization



Nonnegative Matrix Factorization



Existence & Uniqueness

- Linear least squares problem $\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b}$ *always* has solution

- Solution is *unique* if, and only if, columns of $\boldsymbol{A}$ are *linearly independent*, i.e., $\mathrm{rank}(\boldsymbol{A}) = n$, where $\boldsymbol{A}$ is $m \times n$

- If $\mathrm{rank}(\boldsymbol{A}) < n$, then $\boldsymbol{A}$ is *rank-deficient*, and solution of linear least squares problem is not unique

- For now, we assume $\boldsymbol{A}$ has full column rank n

Normal Equations

- To minimize squared Euclidean norm of residual vector

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

  take derivative with respect to x and set it to $\mathbf{0}$,

$$2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{A}^T\boldsymbol{b} = \mathbf{0}$$

  which reduces to $n \times n$ linear system of *normal equations*

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

Orthogonality

- Vectors v_1 and v_2 are *orthogonal* if their inner product is zero, $v_1^T v_2 = 0$

Orthogonality

- Vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are *orthogonal* if their inner product is zero, $\boldsymbol{v}_1^T \boldsymbol{v}_2 = 0$

- If $m > n$, $\boldsymbol{b}$ generally does not lie in span$(\boldsymbol{A})$, so there is no exact solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$

- Vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ in span$(\boldsymbol{A})$ closest to $\boldsymbol{b}$ in 2-norm occurs when residual $\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$ is *orthogonal* to span$(\boldsymbol{A})$,

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

  again giving system of *normal equations*

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

Orthogonality

- Geometric relationships among $\boldsymbol{b}$, $\boldsymbol{r}$, and span$(\boldsymbol{A})$ are shown in diagram