Designing linear systems — §4.1.1 – §4.1.3

§4.1.1 Regression:

- given collected data pairs $(x^{(k)}, y^{(k)})$, $k = 1, 2, \cdots, m$, $x^{(k)} \in \mathbb{R}^n$
- pick a proper regression function/model $f(x) : \mathbb{R}^n \to \mathbb{R}$ with unknown parameters
- predict $f(x)$ for a new x

Specifically, Let $f_1(x), f_2(x), \cdots, f_n(x)$ be the basis functions, for example, $f_1(x) = 1$, $f_2(x) = x$, $f_3(x) = x^2$, $\cdots$ polynomial.

Suppose

$$f(x) = a_1 f_1(x) + a_2 f_2(x) + \cdots + a_n f_n(x)$$

Then we can use the collected data pairs $(x^{(k)}, y^{(k)})$ to ~~estimate~~ compute the parameters $a_1, a_2, \cdots, a_n$ to determined $f(x)$.

Namely,

$$\begin{cases} a_1 f_1(x^{(1)}) + \cdots + a_n f_n(x^{(1)}) = f(x^{(1)}) = y^{(1)} \\ \vdots \\ a_1 f_1(x^{(m)}) + \cdots + a_n f_n(x^{(m)}) = f(x^{(m)}) = y^{(m)} \end{cases}$$

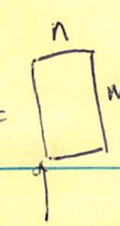
Then we need to solve the linear system

$$Ax = b \qquad (*)$$

where

$$A = \underbrace{\begin{pmatrix} f_1(x^{(1)}) & \cdots & f_n(x^{(1)}) \\ \vdots & & \\ f_1(x^{(m)}) & \cdots & f_n(x^{(m)}) \end{pmatrix}}_{m \times n}, \quad x = \begin{pmatrix} a_1 \\ \vdots \\ \vdots \\ a_n \end{pmatrix}, \quad b = \begin{pmatrix} y^{(1)} \\ \vdots \\ \vdots \\ y^{(m)} \end{pmatrix}$$

Therefore, if $m = n$, we need to solve the square system $(*)$

$A =$ 

§4.1.2 Least squares

If the linear system (*) is overdetermined, i.e, $m > n$ namely we have more data pairs than the number of parameters, then the linear system (*) may not have a solution — see the discussion on solvability.

In this case, we will consider finding a solution x that minimize the difference $\underline{Ax - b}$.
(residual)

Mathematically, Find x such that

$$\min_x \|Ax - b\|_2^2$$

this problem is called a <u>least squares</u> problem

Let $g(x) = \|Ax - b\|_2^2 = (Ax-b)^T(Ax-b)$

$$= x^T A^T A x - 2b^T A x + b^T b$$

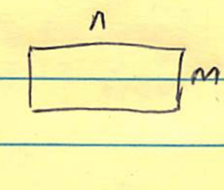
Then the gradient

$$\nabla_x g(x) = 2A^T A x - A^T x$$

therefore, the optimal/solution x is given by the following so-called normal equation:

$$\underline{A^T A}\, x = A^T b$$

Note that the coefficient matrix $\underline{A^T A}$ is $n \times n$ square!

$A =$ 

§4.1.3 Tikhonov regularization

If the linear system (*) is underdetermined, i.e., $m < n$ namely, we have fewer data pairs. Then the linear system (*) have multiple solution — see the discussion on solvability.

In this case, ~~for~~ we select a fixed scalar $\alpha > 0$, solve the Tikhonov regularization problem

$$\min_x \left(\|Ax-b\|_2^2 + \alpha \|x\|_2^2 \right)$$

~~As~~ if $0 < \alpha << 1$, we emphasize $\min \|Ax-b\|_2^2$

as $\alpha \to \infty$, we priorite: $\|x\|_2^2$ = length of x.

Let ~~$g(x)$~~ $g(x) = \|Ax-b\|_2^2 + \alpha \|x\|_2^2$

Then the gradient $\nabla_x g(x) = 2A^TAx - 2A^Tb + 2\alpha x = 0$

Therefore, the optimal x is given by the linear system:

$$(A^TA + \alpha I)x = A^Tb.$$

Key question: How to choose an optimal α? → beyond the scope for our study now!

§4.2.1. Special linear systems:

Symmetric positive definite matrices (SPD) and the Cholesky decomposition.

Definition. A matrix $B \in \mathbb{R}^{n\times n}$ is sym. pos. def. if

(1) $B^T = B$

(2) $x^T B x > 0$ whenever $x \neq 0$

Denote: $B > 0$

Example: if cols of A are linearly independent, then $A^T A$ is SPD.

verify: symmetry: $(A^T A)^T = A^T (A^T)^T = A^T A$

pos. def. $x^T A^T A x = \|Ax\|_2^2 > 0$

since $x \neq 0$ and columns of A are linearly independent.

Gaussian Elim. (~~to~~ LU factorization) to ~~the~~ a SPD matrix

Let $C \in \mathbb{R}^{n\times n}$ is SPD and write

$$C = \begin{array}{c} \left(\begin{array}{c|c} c_{11} & v^T \\ \hline v & \tilde{C} \end{array} \right) \begin{array}{l} 1 \\ n-1 \end{array} \\ \begin{array}{cc} 1 & \quad n-1 \end{array} \end{array}$$

Then since $e_1^T C e_1 = c_{11} > 0$, where $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, we can define

$$E_1 = \begin{array}{c} \left(\begin{array}{c|c} 1/\sqrt{c_{11}} & \\ \hline \gamma & I \end{array} \right) \begin{array}{l} 1 \\ n-1 \end{array} \\ \begin{array}{cc} 1 & \quad n-1 \end{array} \end{array}, \quad$$

γ is a vector to be determined for eliminating v in C.

Then
$$E_1 C = \begin{pmatrix} \sqrt{c_{11}} & v^T/\sqrt{c_{11}} \\ c_{11} r + v & \tilde{C} + r v^T \end{pmatrix} = \begin{pmatrix} \sqrt{c_{11}} & v^T/\sqrt{c_{11}} \\ 0 & \tilde{C} + r v^T \end{pmatrix}$$

Choose $r = -\frac{v}{c_{11}}$

and
$$(E_1 C) E_1^T = \begin{pmatrix} \sqrt{c_{11}} & v^T/\sqrt{c_{11}} \\ 0 & \tilde{C} + r v^T \end{pmatrix} \begin{pmatrix} 1/\sqrt{c_{11}} & r^T \\ 0 & I \end{pmatrix}$$
$$= \begin{pmatrix} 1 & \sqrt{c_{11}}\, r^T + v^T/\sqrt{c_{11}} \\ 0 & \tilde{C} + r v^T \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{C} - \frac{1}{c_{11}} v v^T \end{pmatrix} = \left(\begin{array}{c|c} 1 & \\ \hline & D \end{array}\right)$$

since $r = -\frac{v}{c_{11}}$

In addition, one can show $D = \tilde{C} - \frac{1}{c_{11}} v v^T$ is sym. pos. def (Homework prob).

Since D is SPD, we can systematically repeat the above procedure to the matrix C in the same way to the matrix D, to find $E_2, E_3, \cdots E_n$ such that

$$\underline{E_n \cdots E_2 E_1}\, \underline{C}\, \underline{E_1^T E_2^T \cdots E_n^T} = I$$

Consequently:
$$C = \underbrace{(E_n \cdots E_2 E_1)^{-1}}\, \underbrace{(E_1^T E_2^T \cdots E_n^T)^{-1}}$$
$$= L \cdot L^T \longleftarrow \text{Cholesky Factorization}$$

where $L = (E_n \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_n^{-1}$ = low triangular

<u>Example.</u> $$C = \begin{pmatrix} 4 & -2 & 4 \\ -2 & 5 & -4 \\ 4 & -4 & 14 \end{pmatrix}$$

- Let $$E_1 = \begin{pmatrix} 1/2 & 0 & 0 \\ 1/2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}.$$

  Then $$E_1 C E_1^T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 10 \end{pmatrix}$$

- Let $$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 1 \end{pmatrix}$$

  Then $$E_2 \left(E_1 C E_1^T\right) E_2^T = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 9 \end{pmatrix}$$

- Let $$E_3 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1/3 \end{pmatrix}$$

  Then $$E_3 \left(E_2 E_1 C E_1^T E_2^T\right) E_3^T = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$$

- In summary

$$C = \underline{\left(E_1^{-1} E_2^{-1} E_3^{-1}\right)} \left(E_1^{-1} E_2^{-1} E_3^{-1}\right)^T = L \cdot L^T \Leftarrow \text{Cholesky Factorization}$$

where $$L = E_1^{-1} E_2^{-1} E_3^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 2 & -1 & 3 \end{pmatrix}$$