Content

1. Line Fitting: $x \in \mathbb{R}^1$, $y \in \mathbb{R}^1$
2. Vector/Matrix Notation
3. Non-Linear Response: Polynomial Curve Fitting
4. Regularized Least Square
5. ~~From Basis Function Perspective~~

1. Line Fitting: $y = wx + b$ $\quad y = 2x+1$

Data: $D = \{\langle x_i, y_i \rangle\}_{i=1}^{n}$ $x_i \in \mathbb{R}$, $y_i \in \mathbb{R}$, $i = 1, 2, \cdots, n$

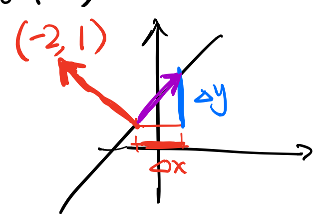
x	-1	0	1	2	3	$\cdots$	x_n
y	-0.9	1.1	2.8	5	7.2		y_n

Model: $y = wx + b$. ($w.b$ 待求)

$$\frac{dy}{dx} = w$$

① $w\,dx - dy = 0$

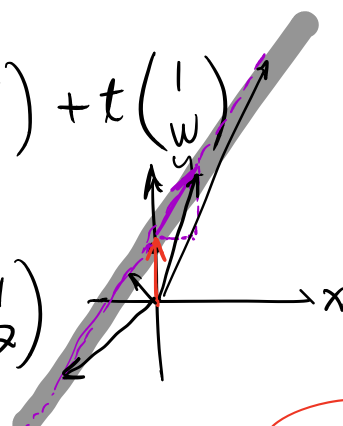
$$(-w \;\; +1)\begin{pmatrix} dx \\ dy \end{pmatrix} = 0$$



② Subspace:
$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix} + t\begin{pmatrix} 1 \\ w \end{pmatrix}$$

$y = 2x+1$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$



Loss

$$\hat{y}_i = w x_i + b$$

$$\mathcal{L}(\hat{y}_i, y_i) = \cancel{\tfrac{1}{2}}(y_i - \hat{y}_i)^2 = \frac{1}{2}(y_i - (wx_i + b))^2$$

Known

Unknown parameter

$$\mathcal{L}_D(w, b) = \sum_{i=1}^{n} \mathcal{L}(y_i, \hat{y}_i)$$

$$= \frac{1}{2}\sum_{i=1}^{n}\Big(y_i - \underbrace{(wx_i + b)}\Big)^2$$

$$= \frac{1}{2}\sum_{i=1}^{n}\Big(y_i^2 + (x_i^2)w^2 + (2x_i)(w\cdot b) + b^2 - (2x_iy_i)w - (2y_i)b\Big)$$

$$= \Big(\overbrace{\frac{1}{2}\sum_{i=1}^{n}x_i^2}^{>0}\Big)w^2 + \Big(\overbrace{\frac{1}{2}n}^{>0}\Big)b^2 + \Big(\sum_{i=1}^{n}x_i\Big)w\cdot b - \Big(\sum_{i=1}^{n}x_iy_i\Big)w - \Big(\sum_{i=1}^{n}y_i\Big)b$$

$$\begin{cases} \dfrac{\partial L}{\partial w} = \Big(\underbrace{\sum_{i=1}^{n}x_i^2}_{n\overline{x^2}}\Big)w - \Big(\underbrace{\sum_{i=1}^{n}x_i}_{n\overline{x}}\Big)b - \Big(\underbrace{\sum_{i=1}^{n}x_iy_i}_{n\overline{xy}}\Big) = 0 \\ \dfrac{\partial L}{\partial b} = \Big(\underbrace{\sum_{i=1}^{n}x_i}_{n\overline{x}}\Big)w + (n)b - \Big(\underbrace{\sum_{i=1}^{n}y_i}_{n\overline{y}}\Big) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \overline{x^2}\,w - \overline{x}\,b = \overline{xy} \\ \overline{x}\,w + b = \overline{y} \end{cases}$$

$$\Rightarrow \begin{cases} w = \dfrac{\begin{vmatrix} -\overline{x} & \overline{xy} \\ 1 & \overline{y} \end{vmatrix}}{\begin{vmatrix} \overline{x^2} & -\overline{x} \\ \overline{x} & 1 \end{vmatrix}} = \dfrac{\overbrace{\overline{xy} - \overline{x}\,\overline{y}}^{\mathrm{Cov}(X,Y)\quad (EXY)-(EXEY)}}{\underbrace{\overline{x^2} - (\overline{x})^2}_{DX = EX^2 - (EX)^2}} \quad (*) \\ b = \overline{y} - w\overline{x} \end{cases}$$

DONE

x	1	2	⋮	$\bar{x}$
y	3	5	⋮	$\bar{y}$
x^2	1	4		$\overline{x^2}$
$x_i y_i$	3	10		$\overline{x_i y_i}$

$$\frac{cov(x, y)}{cov(x, x)}$$

2. Vector / Matrix Notation

$$y = \underbrace{w}_{w_1} x + \underbrace{b}_{w_0} = (\cancel{b\ \ w})_{w_0\ \ w_1} \begin{pmatrix} 1 \\ x \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}_{n \times d} \quad d = 2.$$

$$\vec{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} \qquad x_{i1} = 1, \ \forall i,$$

$$D = \{\langle x_i, y_i \rangle\}_{i=1}^n, \quad x_i \in \mathbb{R}^d, \ d = 2. \quad y_i \in \mathbb{R}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{n \times 1}$$

$$\mathcal{L}_D(w) = \frac{1}{2} \sum_{i=1}^n \left(\underline{y_i - w^T x_i}\right)^2$$

$$= \frac{1}{2} \| Xw - y \|_2^2$$

$$= \frac{1}{2} (Xw - y)^T (Xw - y)$$

$$= \frac{1}{2} \left[w^T X^T X w - w^T X^T y - y^T X w + y^T y \right]$$

$$X\vec{w} = \begin{pmatrix} w_0 + w_1 x_1 \\ w_0 + w_1 x_2 \\ \vdots \\ w_0 + w_1 x_n \end{pmatrix}$$

$$Xw - y = \begin{pmatrix} w_0 + w_1 x_1 - y_1 \\ w_0 + w_1 x_2 - y_2 \end{pmatrix}$$

$$\begin{pmatrix} \vdots \\ w_0 + w_1 x_n - y_n \end{pmatrix}$$

$$\mathcal{L}_D(w) = \frac{1}{2}\left(w^T (X^T X) w - 2 w^T (X^T y) + y^T y\right)$$

$$\frac{\partial \mathcal{L}_D(w)}{\partial w} = (X^T X) w - X^T y = 0$$

$$\Rightarrow w = (X^T X)^{-1} X^T y \quad (*) \quad = \frac{X^T y}{X^T X}$$

作业 验证: $(*)$ 与 $(**)$ 等价!

3. Non-Linear Response

(补充: Taylor 展开式)

$$y = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M$$

$$= (w_0 \quad w_1 \quad \cdots \quad w_M) \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^M \end{pmatrix}$$

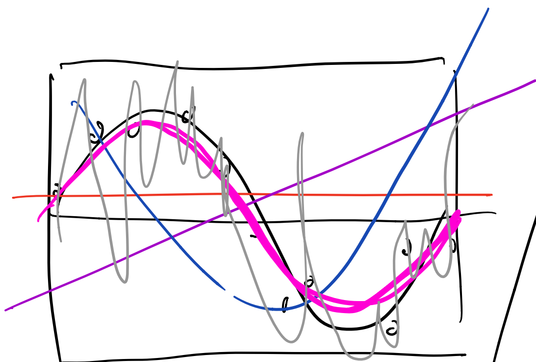
$$= w^T \phi(x)$$

$$\phi(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^M \end{pmatrix}$$

$$\phi_i \in \mathbb{R}^d, \quad d = M + 1$$

$$\Phi = \begin{bmatrix} — \phi_1 — \\ \vdots \\ — \phi_n^T — \end{bmatrix}$$

$M = ?$



$y =$

$$\frac{dy}{dx} = w$$