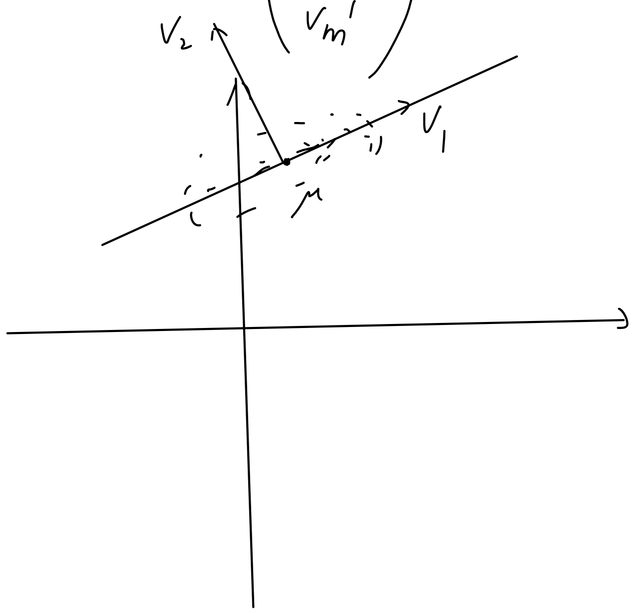
回顾上次：SVD 与 PCA（主成分分析）

$$A = \begin{pmatrix} \alpha_1^T \\ \vdots \\ \alpha_m^T \end{pmatrix} \qquad \alpha_i \in \mathbb{R}^n.$$

m 个 n 维数据.

$$\mu = \frac{1}{m}\sum \alpha_i, \qquad \tilde{A} = A - \begin{pmatrix} \mu^T \\ \vdots \\ \mu^T \end{pmatrix} = QDP^T$$

$$P^T = \begin{pmatrix} v_1^T \\ \vdots \\ v_m^T \end{pmatrix}$$



仿射直线 l: $\mu + v_1 t$ 是

使 $\sum_i \mathrm{dist}(\alpha_i, l)^2$

最小的线.

数据关系的另一种求解形式.

由 n 维数据 $(a_1, \cdots, a_n)$, 预测 y

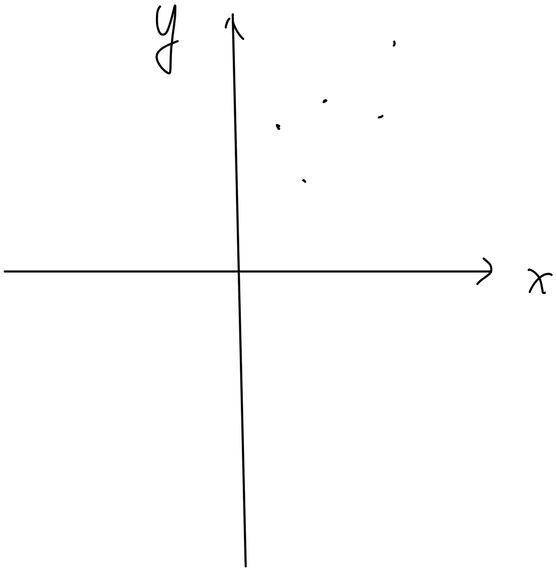
$$y = x_1 a_1 + \cdots + x_n a_n \qquad \text{(线性化模)}$$

多次试验之后，找到一系列方程组

$$\begin{pmatrix}\alpha_1^T\\ \alpha_2^T \\ \vdots \\ \alpha_m^T\end{pmatrix}\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix} = \begin{pmatrix} b_1\\ \vdots \\ b_m\end{pmatrix}$$

(大概率无解)

$m >> n$



$n=1$, $m=5$, 5个数据不在一条直线上.

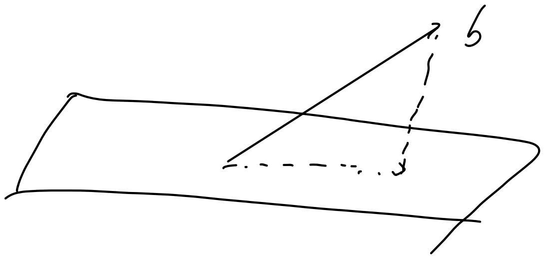
$Ax=b$ $b \notin$ Column space of A.

找最接近的解? 即 $|Ax-b|^2$ 最小.

即使得模型预测的值与 m 次试验结果误差的平方和最小. $\sum_{i=1}^{m}(\alpha_i^T \cdot x - b_i)^2$

$A=(u_1 \cdots u_n)$, $u_i \in \mathbb{R}^m$.

Ax 取遍 Column space of $A = W$



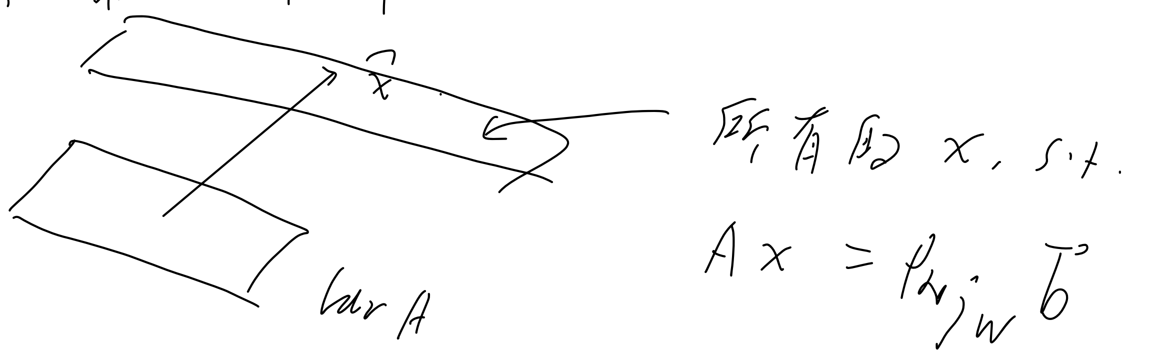
$|b - \mathrm{Proj}_W \vec{b}| =$ distance of b to W.

即 $A\cdot x = \text{Proj}_W \vec{b}$，或者 $b - A\cdot x \perp W$.

另一方面，这样的 x 不唯一确定. 互相之间差 $\ker A$ 的元素

通常会要求 minimize x 的长度来给出唯一解. 即 $x \perp \ker A$

称为最优最小二乘解.



与 SVD 关系.

$$A = QDP^T, \quad \text{令 } A^+ = PD^+Q^T. \ (\text{广义逆})$$

$$D = \begin{bmatrix} \begin{matrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{matrix} & 0 \\ 0 & 0 \end{bmatrix}_{m\times n} \qquad D^+ = \begin{bmatrix} \begin{matrix} \sigma_1^{-1} & & \\ & \ddots & \\ & & \sigma_r^{-1} \end{matrix} & 0 \\ 0 & 0 \end{bmatrix}_{n\times m}$$

定理：对 $Ax = b$. 取 $\hat{x} = A^+ b$ 是最优最小二乘解.

Pf: ① $A\cdot\hat{x} - b \perp \text{Im}\, A$

即 $A^T \cdot (A(A^+ b) - b) = 0$

$$P D^T Q^T \cdot (Q D P^T P D^+ Q^T - I_m) \cdot b = 0$$

$$= (P \underbrace{D^T D D^+}_{D^T} Q^T - P D^T Q^T) \cdot b$$

$$= 0$$

② $\ker A \perp \underbrace{A^+ b}$ $\qquad QD(P^T x) = 0$

$$\ker A = \ker(Q D P^T)$$

$$= P \ker(Q D) = P \cdot \ker D$$

$$A^+ b = P D^+ (Q^T b)$$

即可验证 $D^+ \cdot (Q^T b) \perp \ker D$

$$\ker D = \left\{ \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \hline * \end{pmatrix}_{n\times 1} \right\} \qquad \operatorname{Im} D^+ = \left\{ \begin{pmatrix} * \\ \hline 0 \end{pmatrix}_{n} \right\}$$

$\Rightarrow \ker D \perp \operatorname{Im} D^{+}$.

一种变体模型 $y = x_1 a_1 + \cdots x_n a_n + \underline{c}$

视为 $c \cdot 1$

则 $\underbrace{\begin{pmatrix} \alpha_1^T & 1 \\ \alpha_2^T & 1 \\ \vdots & \vdots \\ \alpha_m^T & 1 \end{pmatrix}}_{\tilde{A}}{}_{m\times(n+1)} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_3 \\ c \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$

$\tilde{A} \cdot \tilde{x} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$

对 $\tilde{A}$ 作 SVD. $\underline{\tilde{A}^{+}}$ 即可