March 3, 2021

$$\underset{n\times n}{A}\;\underset{n\times 1}{x} = \underset{n\times 1}{b}$$

$$\underset{m\times n}{A}\;\underset{n\times 1}{x} \approx \underset{m\times 1}{b}$$

$m > n$

$b_i \approx a_i^T x + c$

$$[A \; \mathbb{1}] \begin{pmatrix} x \\ c \end{pmatrix} \approx b$$

$$\sum_i (b_i - a_i^T x)^2$$

$$\boxed{\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 \quad \text{linear least squares}}$$

Examples:

(1) Data fitting

$(z_1, f(z_1)), \dots, (z_m, f(z_m))$

$f(z) \approx x_3 z^3 + x_2 z^2 + x_1 z + x_0$

$a_i^T = [1 \; z_i \; z_i^2 \; z_i^3] \qquad b_i = f(z_i)$

$b_i \approx a_i^T x \qquad \sum_{i=1}^m (b_i - a_i^T x)^2$

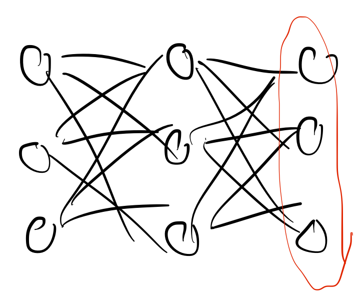
$\|Ax - b\|_2^2 \qquad n = 4$

(2) Stats / ML

a_i = data about page i

b_i = # clicks

$\|Ax - b\|_2^2 \qquad \tilde{a}_i^T x$ (linear regression)



final layer $= a_i$

$$\min_x \|Ax - b\|_2^2$$ Why squared error?

One reason: "right" norm for reasonable stat models

- $b_i = a_i^T x + \varepsilon_i$ $\quad \varepsilon_i \sim N(0, \sigma^2)$ i.i.d. $\quad p(\varepsilon) = \dfrac{\exp\left(-\frac{1}{2}\varepsilon^2/\sigma^2\right)}{\sqrt{2\pi}\,\sigma}$

$$b_i - a_i^T x = \varepsilon_i$$

Given x, likelihood (A, b) is

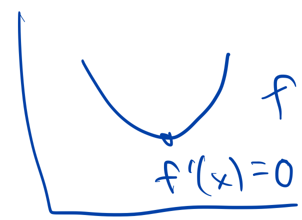
$$\prod_{i=1}^m p(b_i - a_i^T x)$$

$$= \prod_{i=1}^m \exp\left(-\tfrac{1}{2}(b_i - a_i^T x)^2/\sigma^2\right) \Big/ \sqrt{2\pi}\,\sigma$$

$$\tfrac{1}{2} \sum_{i=1}^m (b_i - a_i^T x)^2 / \sigma^2)$$

minimize $\Rightarrow$ maximize likelihood

$$H_{ij} = \frac{\partial \, \partial}{x_i \, x_j}$$



$f''(x) > 0$