Sep 30, 2020

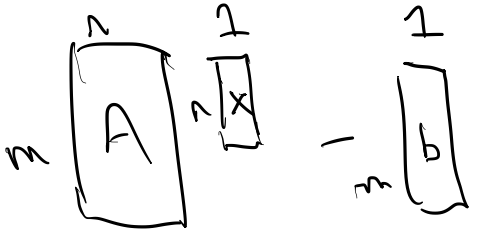
New topic: least squares, QR factorization

Model: $b_i \approx a_i^T x + z \qquad i = 1, \ldots, m$

b_i: outcome; a_i: features; z: intercept

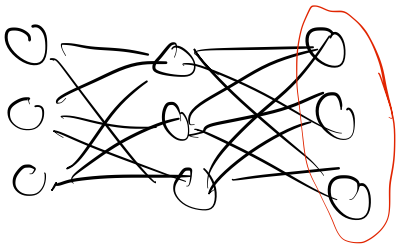
Error: $\sum_{i=1}^{m} (a_i^T x + z - b_i)^2 \Rightarrow \left\| \begin{pmatrix} a_1^T & 1 \\ \vdots & \vdots \\ a_m^T & 1 \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} - b \right\|_2^2$

Linear least squares: $\min_x \|Ax - b\|_2^2$



Examples:

(1) stats/ML a_i = data about page i, b_i = # clicks



final layer $= a_i$

#IJALM

(2) polynomial fitting $a_i^T = (1 \; z_i \; z_i^2 \; z_i^3)$ $b_i = f(z_i)$

$(z_i, f(z_i))$

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

Why squared error?

$\|\cdot\|_1$ $\|\cdot\|_\infty$ ...

$b_i = a_i^T x^*$

1. not unreasonable
2. computationally nice
3. statistical interpretations
4. building block

$= \prod_{i=1}^{m}$

x)

$R\hat{x} = Q^T b$

1. Thin QR $O(mn^2)$
2. $c = Q^T b$
3. $R\hat{x} = c$

How to compute QR?

From what we know...

$$A^TA = R^TQ^TQR = R^TR = LL^T \quad L = R^T$$

(1) Cholesky $\Rightarrow L^T = R$

(2) $A = QR \Rightarrow Q = AR^{-1}$

"Cholesky QR"