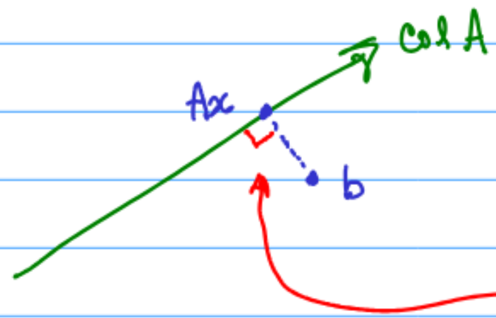
$A \in \mathbb{R}^{m \times n}$ with $m > n$

Idea minimize $\|Ax - b\|$ $\qquad A = QR = \tilde{Q}\tilde{R}$

suppose $b \notin \operatorname{Col} A = \{Ax : x \in \mathbb{R}^n\}$.



closest point Ax in the column space is such that this is perpendicular.

$Ax - b$ needs to be perpendicular to every point in the Col A.

Thus

$(Ax - b) \cdot Ay = 0$ $\quad \boxed{\text{for all } y \in \mathbb{R}^n.}$

Make QR factorization of A

$$\underset{m\times n}{A} = \underset{m\times m}{Q}\,\underset{m\times n}{R} = \underset{m\times n}{\tilde{Q}}\,\underset{n\times n}{\tilde{R}}$$

note if the columns of A are linearly independent then $\tilde{R} \in \mathbb{R}^{n\times n}$ is invertible.

$(Ax - b) \cdot \tilde{Q}\tilde{R}y = 0$

$\tilde{Q}^T(Ax - b) \cdot \tilde{R}y = 0$

Since $\tilde{R}$ is invertible $z = \tilde{R}y$. $y = \tilde{R}^{-1}z$ so if $y \in \mathbb{R}^n$ for all elements of $\mathbb{R}^n$ then $z \in \mathbb{R}^n$ over all elements of $\mathbb{R}^n$.

$\tilde{Q}^T(Ax - b) \cdot z = 0$ $\quad \boxed{\text{for all } z \in \mathbb{R}^n}$

that's why reduced $\tilde{Q}\tilde{R}$ here because $\tilde{R}$ is invertible.

✓ implies

$$\tilde{Q}^T(Ax-b)=0$$

even though I can't solve $Ax=b$ because $b \notin \text{Col}\,A$. It is possible to solve this

$$\tilde{Q}^T A x = \tilde{Q}^T b$$

$A=\tilde{Q}\tilde{R}$

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

is this the identity? YES.

Reason is not $\tilde{Q}^T=\tilde{Q}^{-1}$ because $\tilde{Q}$ is not an orthogonal matrix.

But $\tilde{Q}$ does have orthonormal columns

$$\tilde{Q}=\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$$

where $q_i \in \mathbb{R}^m$

and $q_i \cdot q_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i\neq j \end{cases}$

$$\tilde{Q}^T=\begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix}$$

$$\tilde{Q}^T\tilde{Q}=\begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix}\begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}=\begin{bmatrix} q_1\cdot q_1 & q_1\cdot q_2 & \cdots & q_1\cdot q_n \\ q_2\cdot q_1 & q_2\cdot q_2 & \cdots & q_2\cdot q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n\cdot q_1 & q_n\cdot q_2 & \cdots & q_n\cdot q_n \end{bmatrix}=I_{n\times n}$$

note that $\tilde{Q}\tilde{Q}^T$ is a projection onto a n dimensional subspace — not the identity.

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

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

$\tilde{R}$ is upper triangular and invertible.

So can solve for x using substitution....

and the x one gets minimizes $\|Ax-b\|$.

Application least squares fitting of data...

Suppose some data:

x	y
x_1	y_1
x_2	y_2
x_3	y_3
$\vdots$	$\vdots$
x_m	y_m

Fit a linear model to the data...

$$F_c(x) = c_1\phi_1(x) + c_2\phi_2(x) + \cdots + c_n\phi_n(x)$$

$c \in \mathbb{R}^n$

linear as a function of c.

In particular the $\phi_i(x)$ could be any functions you like

$$F_c(x) = c_1 \sin x + c_2 \cos x.$$

$$F_c(x) = c_1 \cdot 1 + c_2 x + c_3 x^2 + c_4 x^3$$

Goal: Find the parameters $c \in \mathbb{R}^n$ for which the data was most likely to have come from.

$$y_i \approx F_c(x_i) + \text{noise}.$$

Assume the noise is given by independent normal random variables...

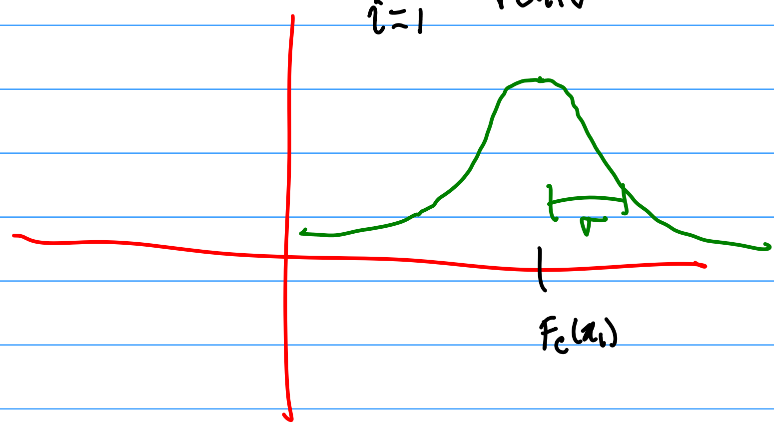
$$Y_i = \underbrace{F_c(x_i)}_{\text{mean}} + \underbrace{\sigma}_{\text{variance}} \eta_i$$

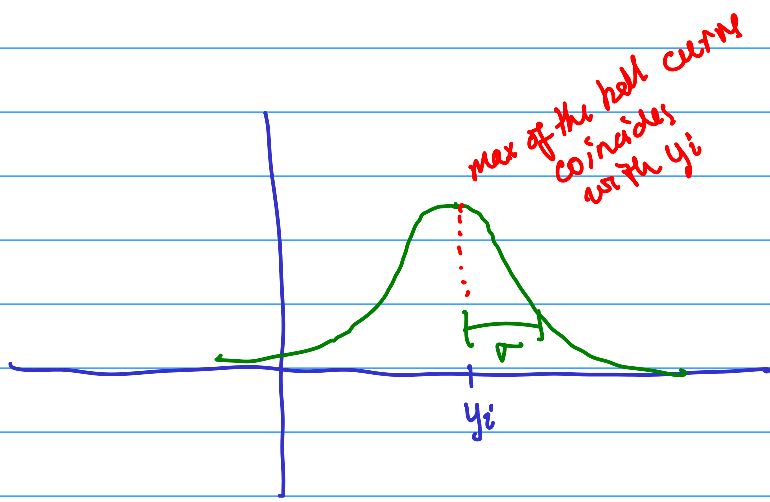
where η_i is an independent standard normal variable.

View y_i as a sample of the random variable Y_i.

$$\text{pdf } Y_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s - F_c(x_i))^2}{2\sigma^2}}$$

$$\text{joint pdf } Y = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s_i - F_c(x_i))^2}{2\sigma^2}}$$





For the joint probability distribution we have to find the parameter c that maximizes.

$$\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - F_c(x_i))^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^m e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - F_c(x_i))^2}$$

because of the negative sign maximizing this is the

same as minimizing $\sum_{i=1}^{m}(y_i - F_c(x_i))^2$. That's least squares.

What's A and what's b.
and what's x so that
minimizing $\|Ax - b\|$ is the same thing...

What is A? Vandermond matrix..

$$F_c(x_i) = c_1\varphi_1(x_i) + c_2\varphi_2(x_i) + \cdots + c_n\varphi_n(x_i)$$

$$Ac = \begin{bmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdots & \varphi_n(x_n) \\ \vdots & & & \\ \varphi_1(x_m) & \varphi_2(x_m) & & \varphi_n(x_m) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} F_c(x_1) \\ F_c(x_2) \\ \vdots \\ F_c(x_m) \end{bmatrix}$$

Next time we do a computation ... and discuss review for the final.