LECT 4. [CONT.] [LJUNG APP II] (1)

$$y_N = \Phi_N \cdot \theta + \varepsilon_N$$

MEAS — LINEAR PREDICTION MODEL — PREDICTION ERRORS

$$\Phi_N \equiv J = \begin{bmatrix} \varphi(1)^T \\ \vdots \\ \varphi(N)^T \end{bmatrix}$$

LINEAR LEAST SQUARES (LLS) ESTIMATOR:

$$\hat{\theta}_N = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T y_N$$

$$= \arg\min_\theta \frac{1}{2} \| y_N - \Phi_N \theta \|_2^2$$

"PREDICTION ERROR MINIMIZATION (PEM)"

4.2
VARIANT: WEIGHTED LEAST SQUARES (WLS)

GIVE DIFFERENT WEIGHTS DUE TO DIFFERENT a) RELIABILITY b) RELEVANCE

$$\min_\theta \frac{1}{2} \sum_{k=1}^{N} \alpha_k (y(k) - \varphi(k)^T \cdot \theta)^2, \quad Q_N = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_N \end{pmatrix}$$

$$= \min_\theta \frac{1}{2} (y_N - \Phi_N \theta)^T Q_N (y_N - \Phi_N \theta)$$

(ALSO GENERAL $Q_N \succ 0$ POSSIBLE) "$\succ 0$" POS. DEF.

CAN SHOW THAT

$$\hat{\theta}_N^{WLS} = \left(\Phi_N^T Q_N \Phi_N\right)^{-1} \Phi_N^T Q_N \cdot Y_N$$

4.3 RESIDUALS AND PREDICTION ERRORS

AT OPTIMAL $\hat{\theta}_N$, CAN COMPARE MODEL PREDICTION

$\boxed{\hat{Y}_N = \Phi_N \hat{\theta}_N}$ AND MEASUREMENTS $\boxed{Y_N}$

ONE DEFINES "R-SQUARE" AS:

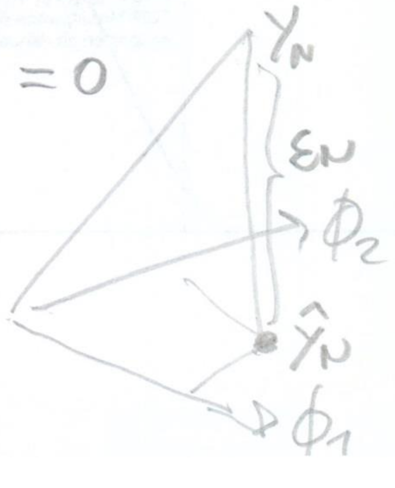
$$R_Y^2 = \frac{\|\hat{Y}_N\|_2^2}{\|Y_N\|_2^2} = \frac{\|Y_N\|_2^2 - \|Y_N - \Phi_N \hat{\theta}_N\|_2^2}{\|Y_N\|_2^2} = 1 - \frac{\|\varepsilon_N\|_2^2}{\|Y_N\|_2^2}$$

(OFTEN, AVERAGE IS SUBTRACTED FROM BOTH FIRST)

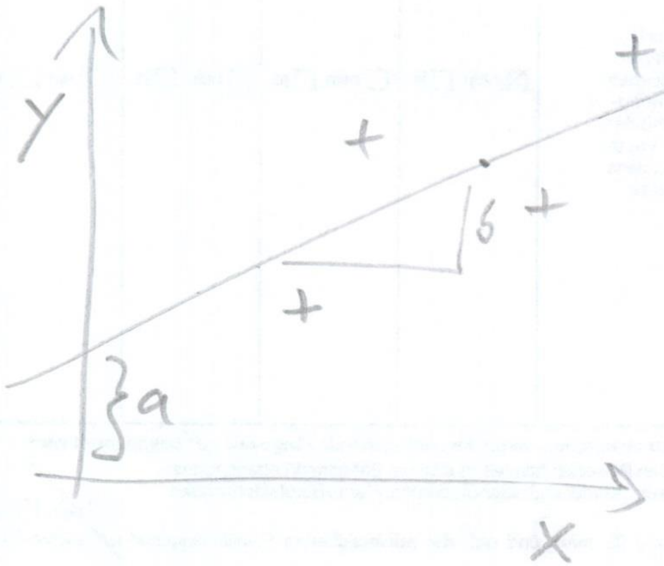
"HOW MUCH OF THE VARIATION IN THE MEASUREMENTS Y IS EXPLAINED BY THE (OPTIMAL) MODEL $\Phi_N \cdot \hat{\theta}_N$"

NOTE:

$$\varepsilon_N^T \hat{Y}_N = (Y_N - \Phi_N \hat{\theta})^T \Phi \hat{\theta} = 0$$



EXAMPLE: FITTING A LINE



MEASUREMENTS
$y(1), \dots, y(N)$
$x(1), \dots, x(N)$

MODEL: $y(k) = a + b\,x(k) + \varepsilon(k)$

REGRESSION VECTOR: $\varphi(k) = \begin{bmatrix} 1 \\ x(k) \end{bmatrix}$

PARAMETER VECTOR: $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$

$$\Phi_N = \begin{bmatrix} 1 & x(1) \\ \vdots & \\ 1 & x(N) \end{bmatrix}, \quad Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}$$

$$\Phi_N \cdot \theta = \begin{bmatrix} 1 & x(1) \\ & \\ & \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a + x(1)\cdot b \\ a + x(2)\cdot b \\ \\ \end{bmatrix}$$

$$\hat{\theta}_N = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N \qquad \hat{\theta}_N = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}$$

$$\hat{Y}_N = \Phi_N \cdot \hat{\theta}_N = \begin{bmatrix} 1 \cdot \hat{a} + x(1) \cdot \hat{b} \\ \vdots \end{bmatrix}$$

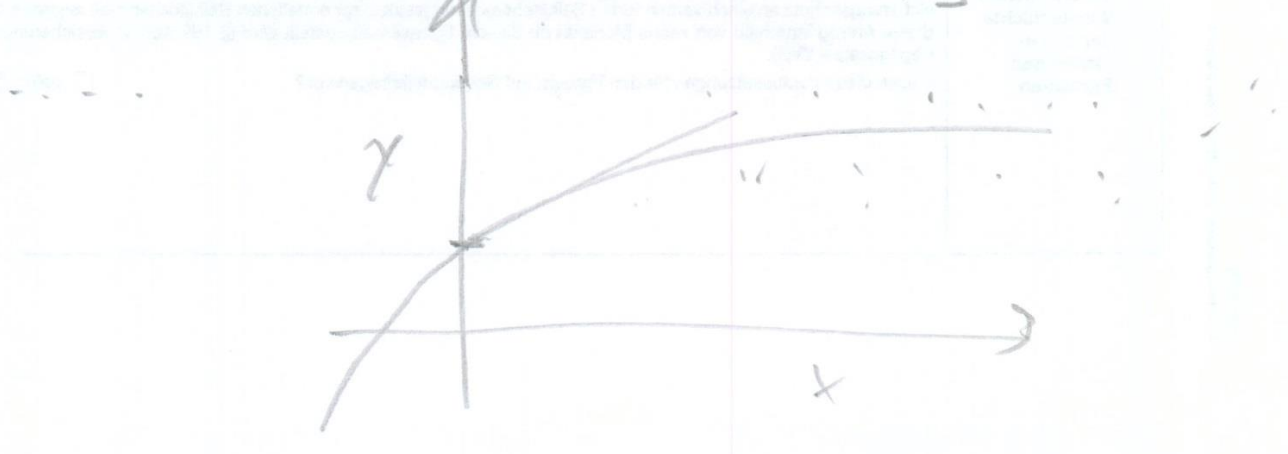
$$\hat{y}(k) = \hat{a} + \hat{b} \cdot x(k)$$

$$R_y^2 = \frac{\|\hat{y}\|_2^2}{\|y\|_2^2}$$

EX 2: FITTING A PARABOLA

$$y(k) = a + b\,x(k) + c \cdot x(k)^2 \qquad \theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

REGR. VECTOR $\varphi(k) = \begin{bmatrix} 1 \\ x(k) \\ x(k)^2 \end{bmatrix}$



ANALYSIS OF LLS

ASS.: • REGRESSORS DETERMINISTIC

• $y(k) = \varphi(k)^T \cdot \theta_0 + w_0(k)$ $\quad W_N = \begin{pmatrix} w_0(1) \\ \vdots \\ w(N) \end{pmatrix}$

MEASUREMENT TRUE PARAMETER

i.i.d. NOISE WITH VARIANCE σ_0, MEAN ZERO

(DROP INDEX N)

$$\hat{\theta}_N = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N$$

$$= (\Phi_N^T \Phi_N)^{-1} \Phi_N^T (\Phi_N \theta_0 + W_N) = \theta_0 + \Phi_N^+ \cdot W_N$$

5.1 BIAS AND COVARIANCE

$\mathbb{E}\, \hat{\theta}_N = \theta_0$ "UNBIASED"

DEF: AN ESTIMATOR $\hat{\theta}_N$ IS UNBIASED IF ITS EXPECTED VALUE IS EQUAL TO THE TRUE VALUE (IF APPLIED TO MEASUREMENTS CORRESPONDING TO TRUE VALUE)

VARIANCE

$$\mathbb{E}\{(\hat{\theta}_N-\theta_0)(\hat{\theta}_N-\theta_0)^T\}$$

$$= \mathbb{E}(\Phi^+ w_N w_N^T \Phi^{+T})$$

$$= \Phi^+ \underbrace{\mathbb{E}\{w_N w_N^T\}}_{\lambda_0 \cdot \mathbb{1}} \Phi^{+T}$$

$$= (\Phi^T\Phi)^{-1}(\Phi^T \Phi)(\phi^T\phi)^{-1} \cdot \lambda_0$$

$$= \lambda_0 \cdot (\phi^T\Phi)^{-1}$$

FOR LARGE N?

$$\Phi_N^T \Phi_N = \sum_{k=1}^{N} \varphi(k)\varphi(k)^T$$

ASSUME: $\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi(k)^T$

$= R_\varphi$ INVERTIBLE

$$\mathrm{COV}(\hat{\theta}_N) \approx \lambda_0 \cdot (R_\varphi \cdot N)^{-1}$$

$$= \frac{\lambda}{N} R_\varphi^{-1}$$

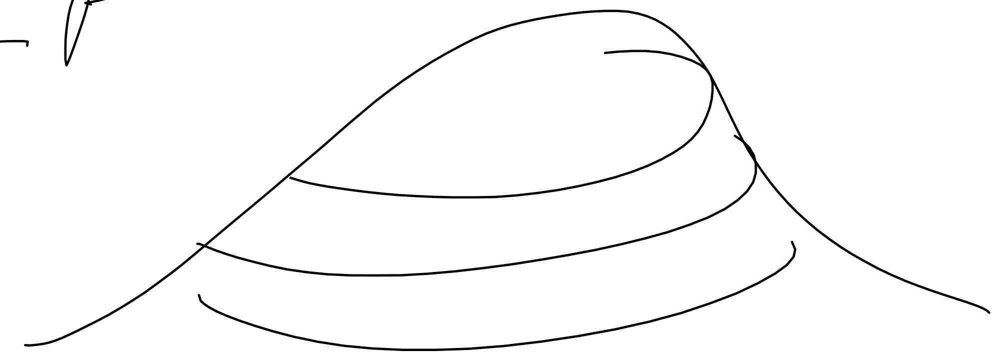
GETS SMALLER FOR N BIGGER

5.2 CONSISTENCY

IN FACT

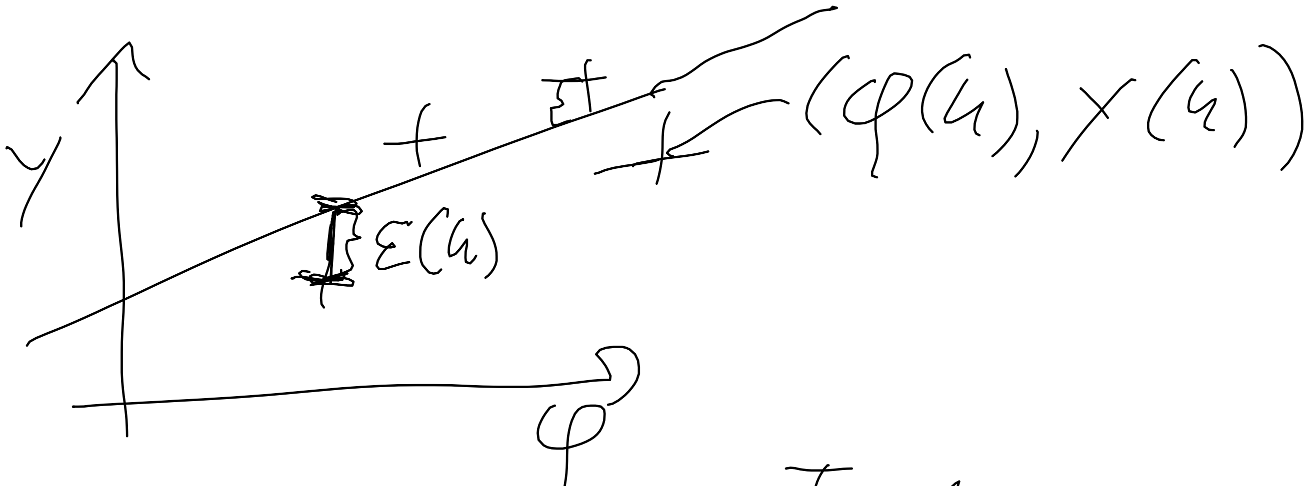
$$\hat{\theta}_N - \theta_0 = \left[\frac{1}{N}\sum \varphi\varphi^T\right]^{-1} \left[\frac{1}{N}\sum \varphi(k)\cdot v(k)\right]$$

$$\to R_\varphi^{-1} \underbrace{\lim \frac{1}{N} \sum \varphi(k)\, v(k)}_{=0 \quad \text{w.p. } 1}$$



pdf $\quad f_\xi(x) = \text{const} \cdot \exp\left(- \dfrac{(x-m)^T P^{-1} (x-m)}{2} \right)$

LINEAR LEAST SQUARES (LLS)



$$y(k) = \varphi(k)^T \cdot \theta + \varepsilon(k) \qquad k = 1, \ldots, N$$

$$y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \qquad J = \Phi_N = \begin{bmatrix} \varphi(1)^T \\ \vdots \\ \varphi(N)^T \end{bmatrix} \qquad \varepsilon_N = \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(N) \end{bmatrix}$$

LLS

$$\hat{\theta}_N = \arg \min_{\theta} \frac{1}{2} \| \underbrace{y_N - \Phi_N \theta}_{= \varepsilon_N} \|_2^2$$

"PREDICTION ERRORS" (PEM).

$$\hat{\theta}_N = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y_N$$
$$= \Phi_N^+ Y_N$$

"PSEUDO INVERSE"
pinv (Φ_N)

[LJUNG APP II]

4.2 WEIGHTED LEAST SQUARES (WLS)

GIVE DIFF. WEIGHT: • DIFFERENT RELIABILITY
" " RELEVANCE

$$\hat{\theta}_N^{WLS} = \arg\min_{\theta} \frac{1}{2} \sum_{k=1}^{N} \alpha(k) \underline{(y(k) - \varphi(k)^T \theta)^2}$$

$\alpha(k) > 0$

$$= \arg\min_{\theta} \frac{1}{2} (Y_N - \Phi_N \theta)^T Q_N (Y_N - \Phi_N \theta)$$

$$Q_N = \begin{pmatrix} \alpha(1) & & 0 \\ & \ddots & \\ 0 & & \alpha(N) \end{pmatrix}$$

$$\hat{\theta}_N^{WLS'} = \left(\Phi_N^T Q_N \Phi_N\right)^{-1} \Phi_N^T Q_N Y_N$$

NOTE: GENERAL Q_N POSSIBLE

$$\boxed{\begin{array}{c} Q_N = Q_N^T \\ eigs(Q_N) > 0 \end{array}}$$

Q_N "POSITIVE DEFINITE"

$$\boxed{Q_N \succ 0}$$

$$Q_N = Q_N^{\frac{1}{2}} \cdot Q_N^{\frac{1}{2}} \quad \text{e.g.}$$

$$Q_N^{\frac{1}{2}} = \begin{pmatrix} \sqrt{a(1)} & & \\ & \ddots & \\ & & \sqrt{a(N)} \end{pmatrix}$$

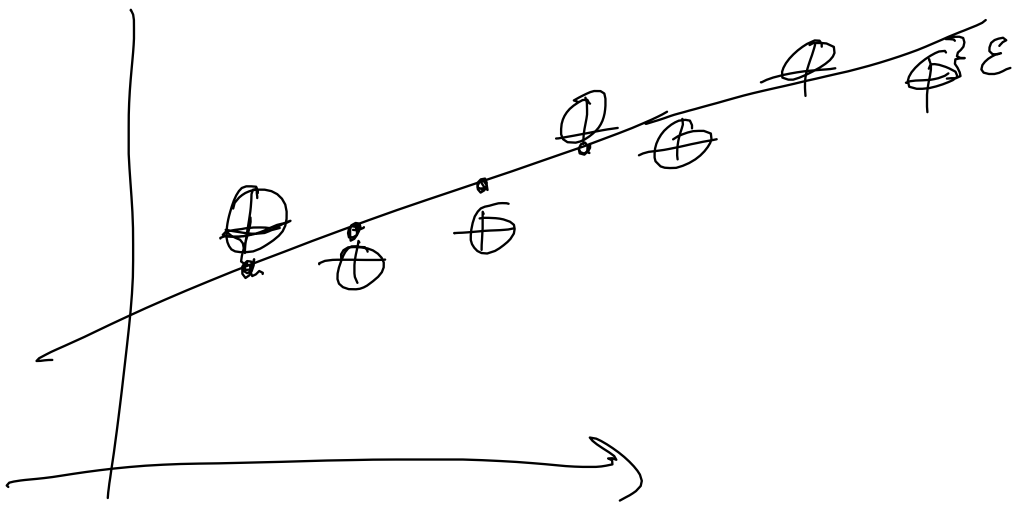
$$\left\| Q_N^{\frac{1}{2}}\left(Y_N - \Phi_N \theta\right)\right\|_2^2$$

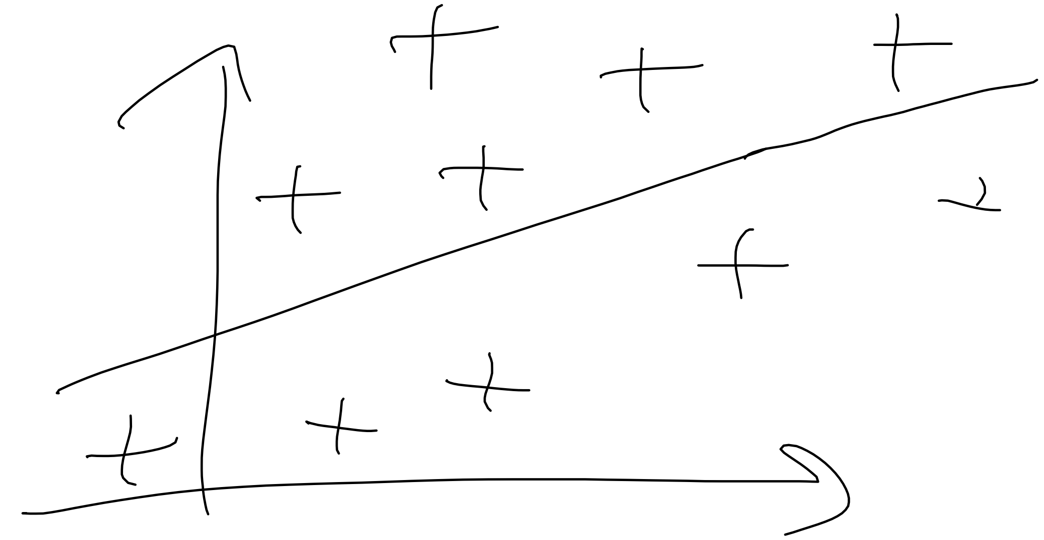
$$= \left(Y_N - \Phi_N \theta\right)^T Q_N \left(Y_N - \Phi_N \theta\right)$$

$$= \Big\| \underbrace{Q_N^{\frac{1}{2}} Y_N}_{=\tilde{Y}_N} - \underbrace{Q_N^{\frac{1}{2}} \Phi_N}_{=\tilde{\Phi}_N} \theta \Big\|_2^2$$

WLS CAN BE REDUCED TO LLS

4.3 RESIDUALS AND PREDICTION ERRORS





COMPARE:

MODEL PREDICTION

$$\hat{y}_N = \Phi_N \cdot \hat{\theta}_N$$

AND

MEASUREMENTS

$$y_N$$

PREDICTION ERROR:

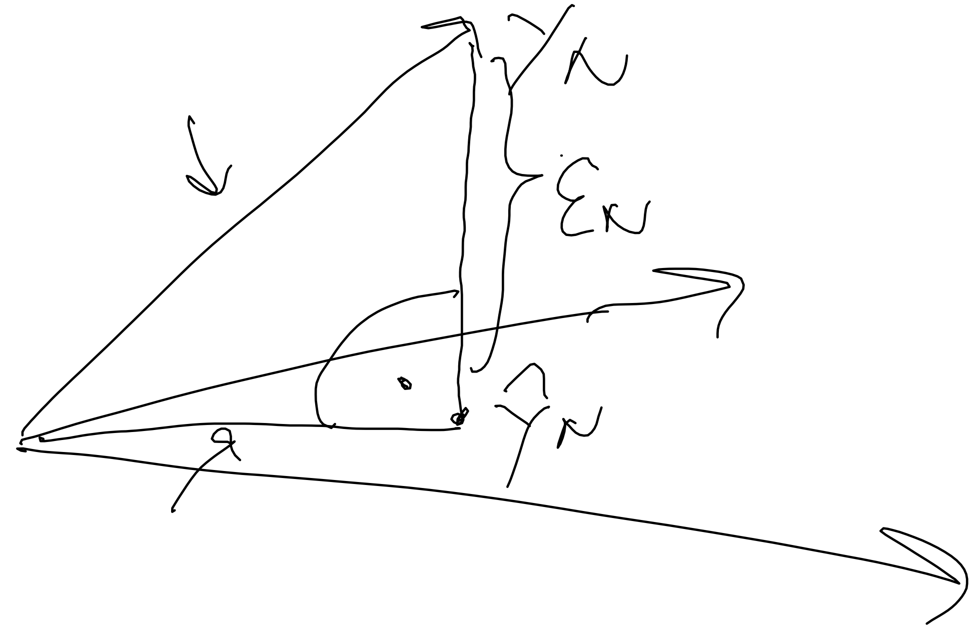
$$\varepsilon_N = y_N - \hat{y}_N$$

OBSERVATION:

$$\varepsilon_N^T \hat{y}_N = (y_N - \Phi_N \hat{\theta}_N)^T \Phi_N \cdot \hat{\theta}_N$$

$$= 0$$

(BEC. OPTIMALITY OF $\hat{\theta}_N$)



$$\|Y_N\|_2^2 = \|\mathcal{E}_N\|_2^2 + \|\hat{Y}_N\|_2^2$$

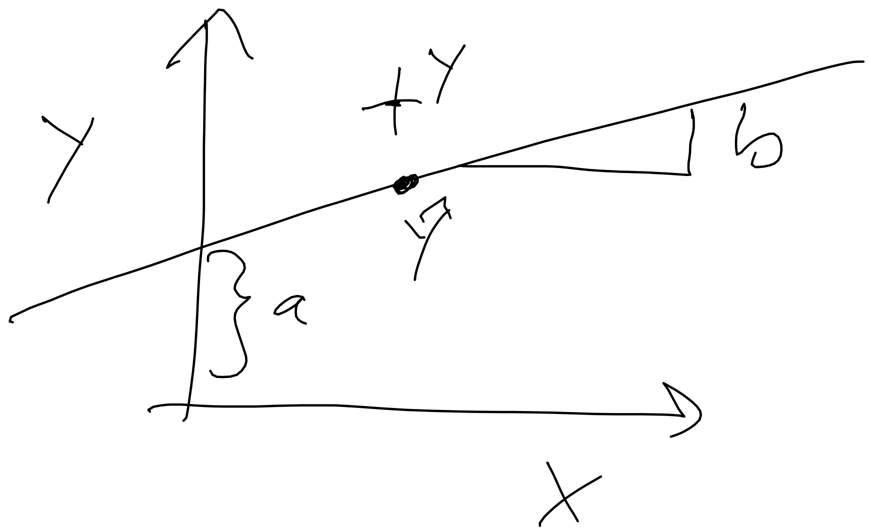
DEF: "R-SQUARE"

$$R_Y^2 = \frac{\|\hat{Y}_N\|_2^2}{\|Y_N\|_2^2} = \frac{\|Y_N\|_2^2 - \|\mathcal{E}_N\|_2^2}{\|Y_N\|_2^2} = 1 - \frac{\|\mathcal{E}_N\|_2^2}{\|Y_N\|_2^2}$$

$$R_Y^2 \in [0, 1] \qquad \text{BIG} = \text{GOOD FIT}$$

"HOW MUCH OF VARIATION CAN BE EXPL.
BY THE MODEL"

EX. FITTING A LINE



$+ (x(k), y(k))$

MEASUREMENTS:

$y(1) \dots y(N)$

$x(1) \dots x(N)$

$$\hat{Y}_N = \Phi_N \cdot \hat{\theta}_N = \begin{bmatrix} \hat{a} + x(1)\hat{b} \\ \vdots \\ \hat{a} + x(N)\hat{b} \end{bmatrix}$$

MODEL: $\boxed{y(k) = a + b\,x(k) + \varepsilon(k)} = \varphi(k)^T \theta + \varepsilon(k)$

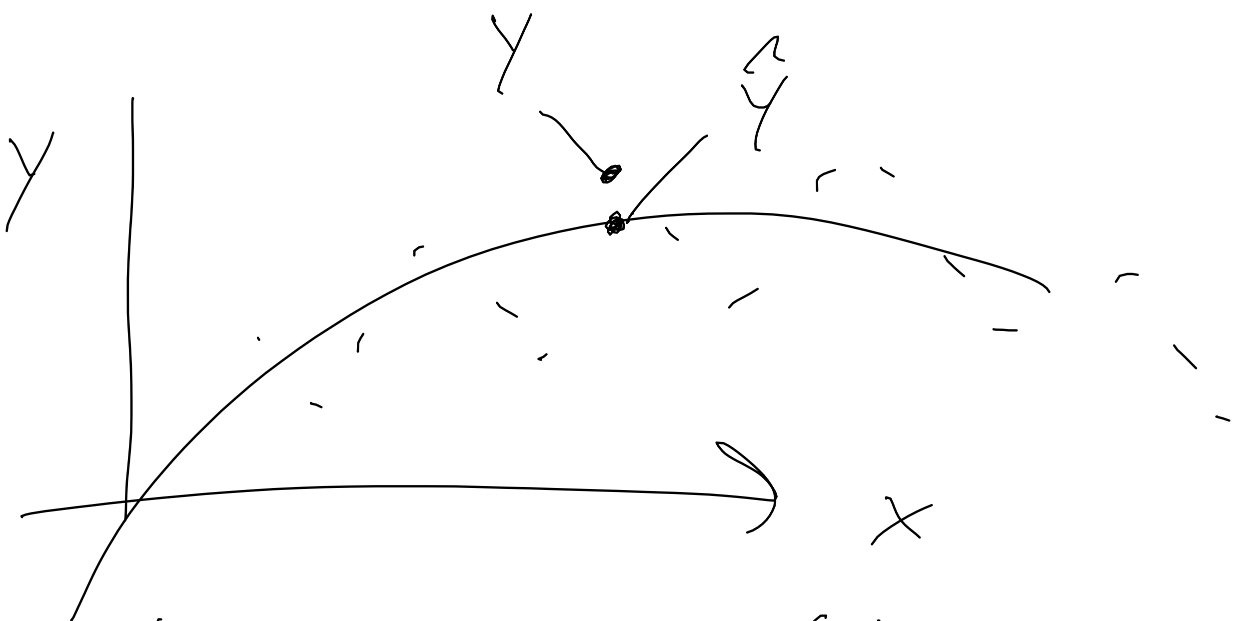
REGRESSION VECTOR: $\varphi(k) = \begin{bmatrix} 1 \\ x(k) \end{bmatrix}$

PARAMETER VECTOR: $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$

$$\Phi_N = \begin{bmatrix} \varphi(1)^T \\ \vdots \\ \varphi(N)^T \end{bmatrix} = \begin{bmatrix} 1 & x(1) \\ 1 & x(2) \\ \vdots & \vdots \\ 1 & x(N) \end{bmatrix} \in \mathbb{R}^{N \times 2}, \quad Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}$$

$$\hat{\theta}_N = (\Phi_N^T \Phi_N)^{-1} \cdot \Phi_N^T \cdot Y_N = \Phi_N^+ \cdot Y_N$$

EX 2: FITTING A PARABOLA



MODEL: $y(k) = a + b\,x(k) + c \cdot x(k)^2 + \varepsilon(k)$

$$\theta = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$\varphi(k)^T = \begin{bmatrix} 1 & x(k) & x(k)^2 \end{bmatrix}$$

$$\Phi_N = \begin{bmatrix} 1 & x(1) & x(1)^2 \\ \vdots & \vdots & \vdots \\ 1 & x(N) & x(N)^2 \end{bmatrix}$$

CRUCIAL: θ ENTER MODEL LINEARLY!