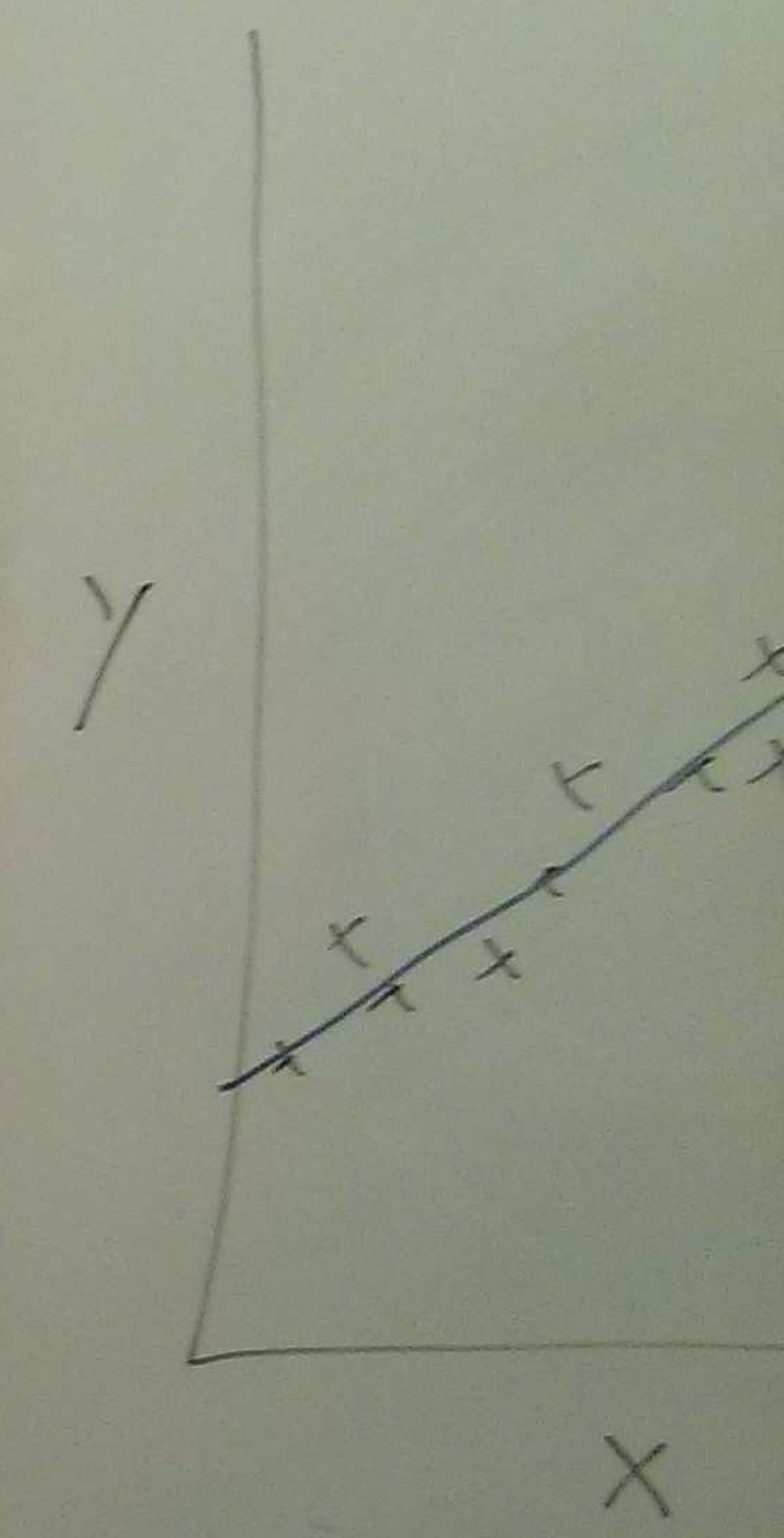
Admin, Linear regression
Robust linear regression
Nonlinear regression
Model Selection
Ridge regression

1. Assignment 2 out
(<u>due Monday</u>)

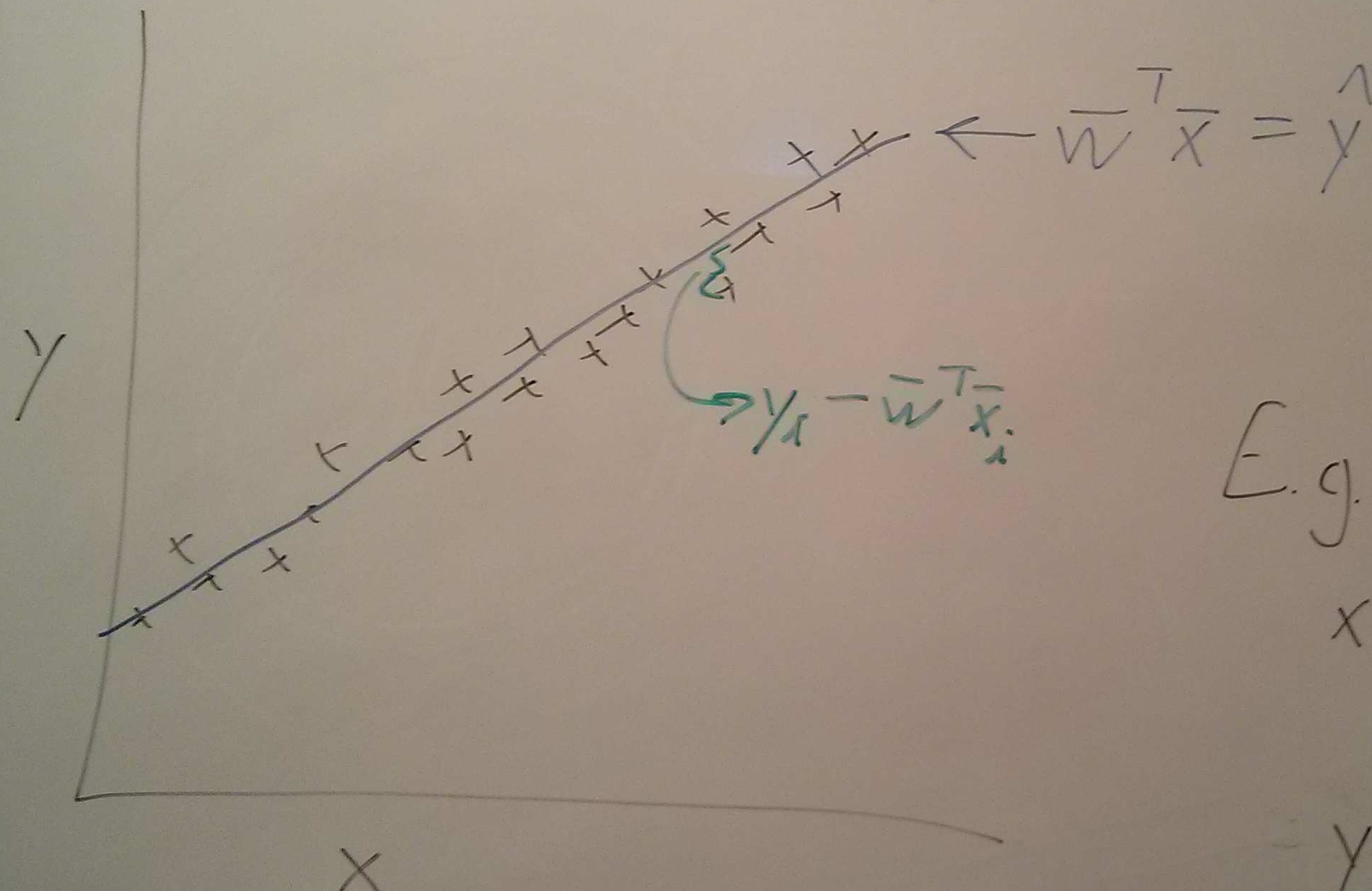
2. Graded assigment 1
also due <u>Monday</u>.
(answers tonight on Piazza)

3. Final "BAD COP" speech
(drop deadline tomorrow)
→ failing is <u>bad</u>
(especially for <u>grad</u> students)

Linear re



Linear regression



E.g.:

x: number of hours spent on ass 2.

y: mark on ass 2.

MLE:

$$\underbrace{\operatorname*{argmin}_{\bar{w} \in \mathbb{R}^d} - \sum_{i=1}^{N} \log p(y_i \mid \bar{x}_i, \bar{w})}_{\text{"negative log-likelihood" (NLL)}}$$

NL

2.

2.

$$NLL(\bar{w}) = \frac{1}{2}(Y - X\bar{w})^T(Y - X\bar{w}) + \text{cons}$$

$$= \frac{1}{2}(Y^T - \bar{w}^T X^T)(Y - X\bar{w})$$

$$= \frac{1}{2}\left[Y^T Y - Y^T X\bar{w} - \bar{w}^T X^T Y + \bar{w}^T X^T X \bar{w}\right]$$

This is a scalar: $[a]^T = [a]$, $\bar{w}^T X^T Y = Y^T X \bar{w}$

$$\nabla NLL(w) = 0 - \bar{w} X^T y + X^T X \bar{w}$$

See assignment 2

Set $\nabla NLL(\bar{w}) = 0$, to get

$$\boxed{X^T X \bar{w} = X^T Y}$$

1. How to s

$X^T X$

Today:

1. How to solve

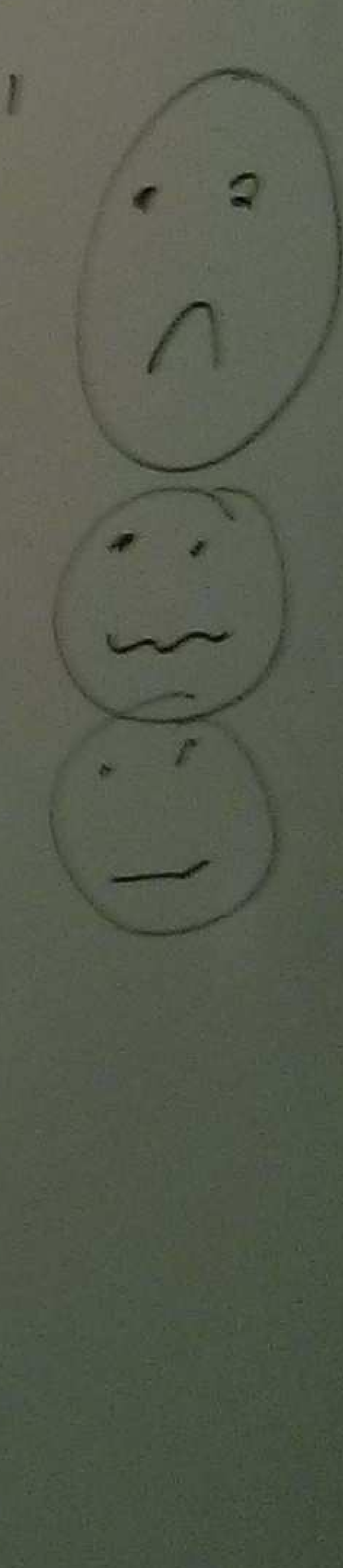
$$\underbrace{X^TX}_{A}\bar{w} = \underbrace{X^TY}_{b}$$

$$A\bar{w} = b$$

$$\bar{w} = (X^TX)^{-1}X^TY$$

"Gauss Jordan"

$O(d^3)$



...lar: $[a]^T = [a]$,

$^TX\bar{w}$

SVD

QR: $\bar{w} = (X^TX)\backslash X^TY$

2. Problems w/ least squares

(and how to fix them)

$y = \bar{w}^T \bar{x}$

Problems w/ least squares

- "non-robust"
- linear assumption
- non-uniqueness

$v \geq |x| = \max\{-x, x\}$

$\downarrow$

$v \geq -x, \; v \geq x$

$-v \leq x \leq v$

L_1 Regression

$$\underset{\bar{w}}{\operatorname{argmin}} \sum_{i=1}^{N}$$

$$\underset{\bar{w}, \bar{v}}{\operatorname{argmin}} \sum_{i=1}^{N} v_i$$

subject to

$v_i \geq | y_i -$

$-v_i \leq y_i -$

L_1 Regression

(Laplace distribution)

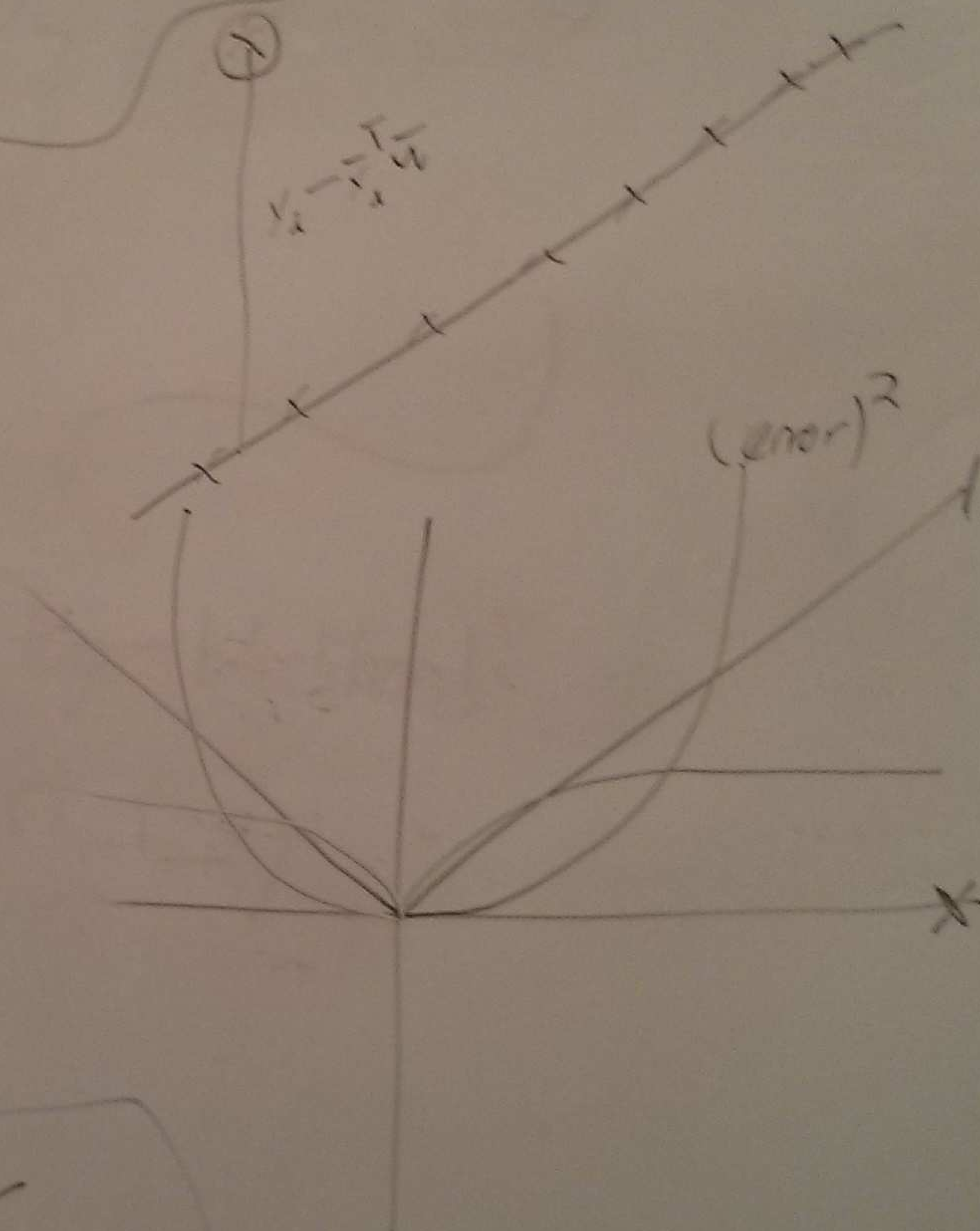
$$\underset{\bar{w}}{\arg\min} \sum_{i=1}^{N} \left| y_i - \bar{x}_i^T \bar{w} \right|$$

$$\boxed{\underset{\bar{w},\bar{v}}{\arg\min} \sum_{i=1}^{N} v_i}$$

subject to

$$v_i \geq \left| y_i - \bar{w}^T \bar{x}_i \right| .$$

$$\boxed{-v_i \leq y_i - \bar{w}^T \bar{x}_i \leq v_i}$$



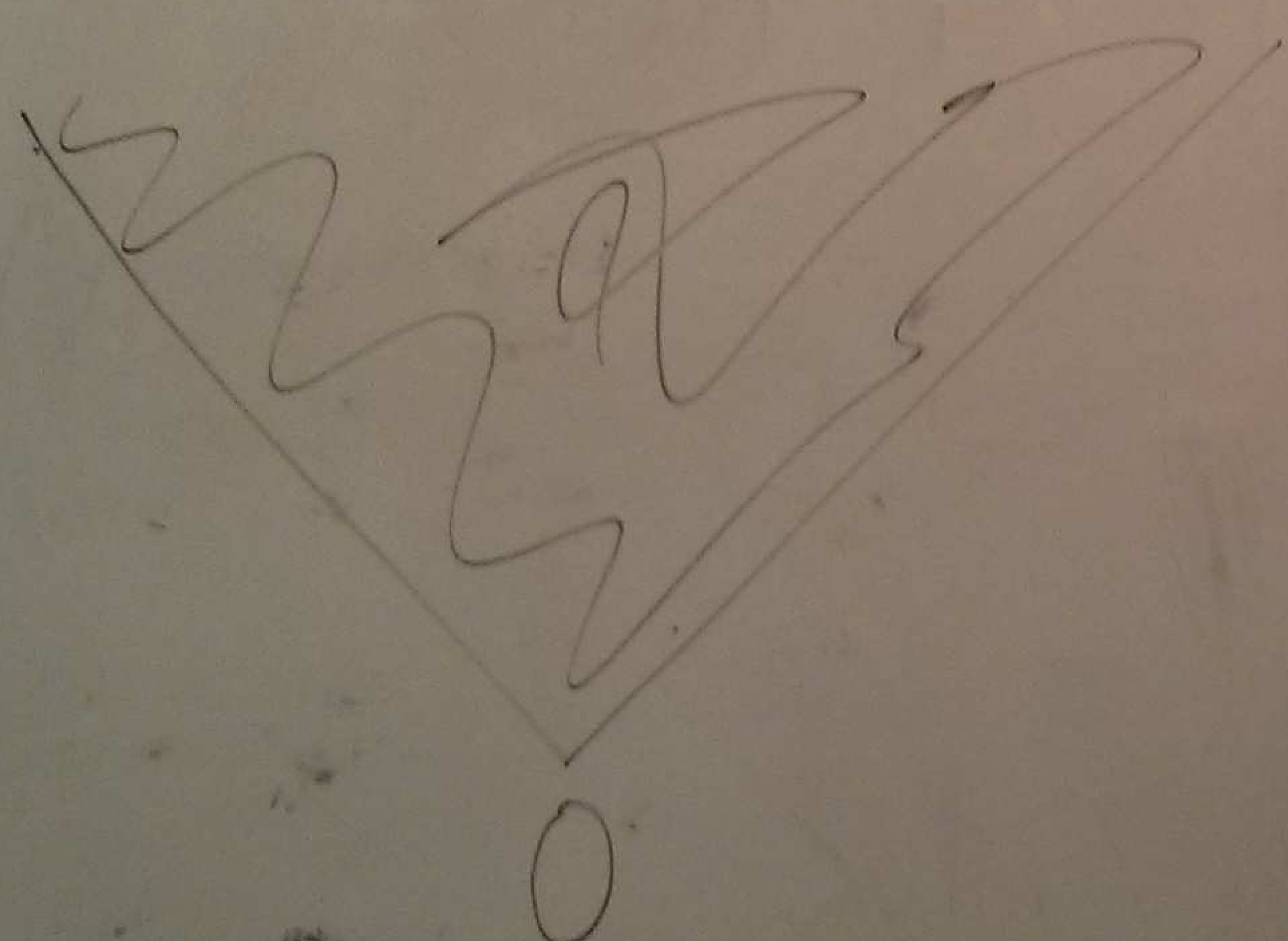
"Linear program"

$$\sum_i |y_i - \bar{x}_i^T \bar{w}| \Longleftrightarrow \sum_i v_i$$

New variable:

$$v_i \geq |y_i - \bar{x}_i^T \bar{w}|, \quad v_i = |y_i - \bar{x}_i^T \bar{w}|,$$

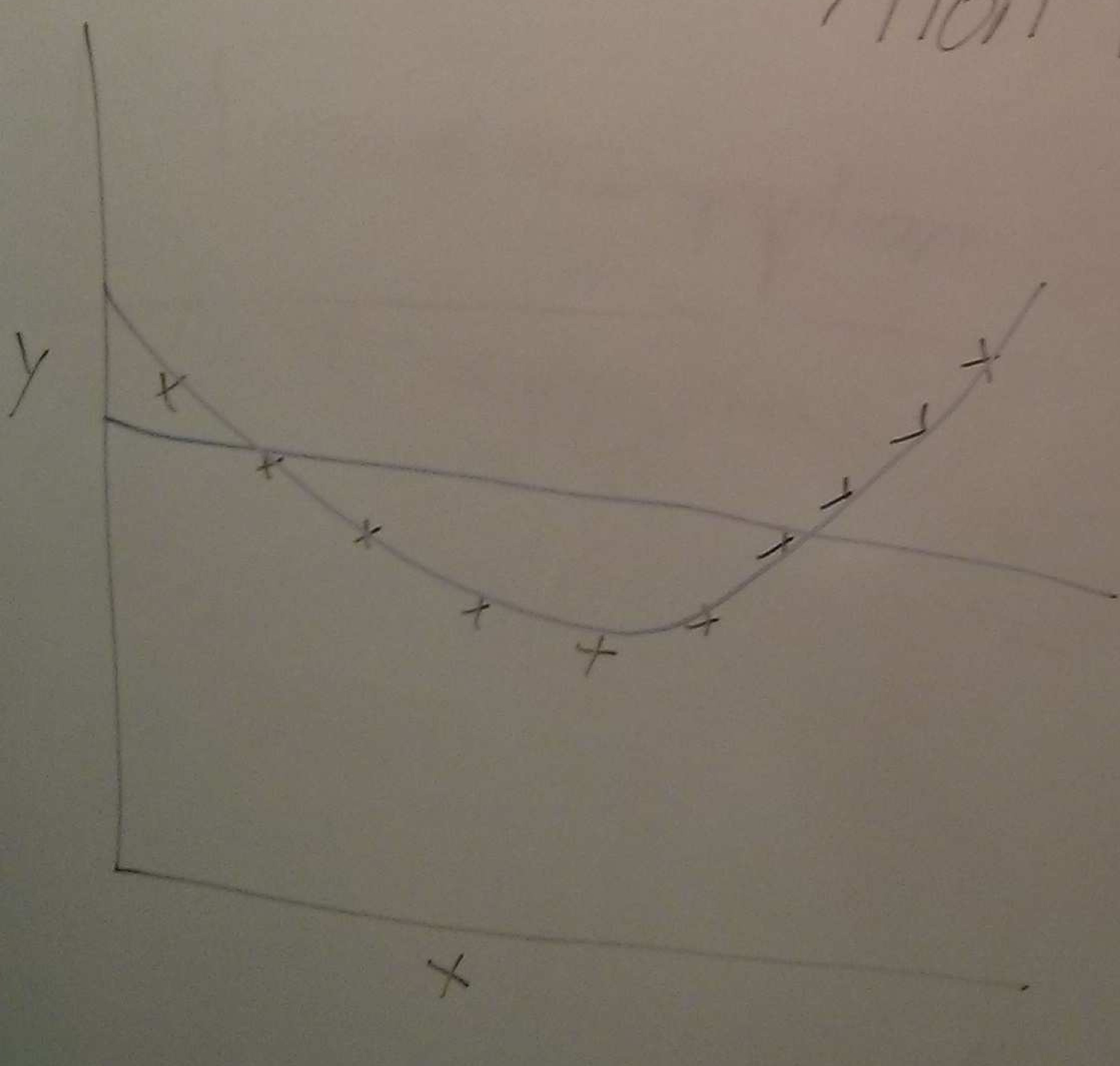
at minimizer.



Basis Expansion

(linear paramet. $(\bar{w})$) + (non-linear tranf. of features $\bar{x}_i$)

$\longrightarrow$ non-linear regression



$y = wx$

$y = w^T [1 \quad x \quad x^2]$

$= w_1 + w_2 x + w_3 x^2$

Bases:

polynomials: $w^T[1, x, x^2, x^3, \dots, x^n]$

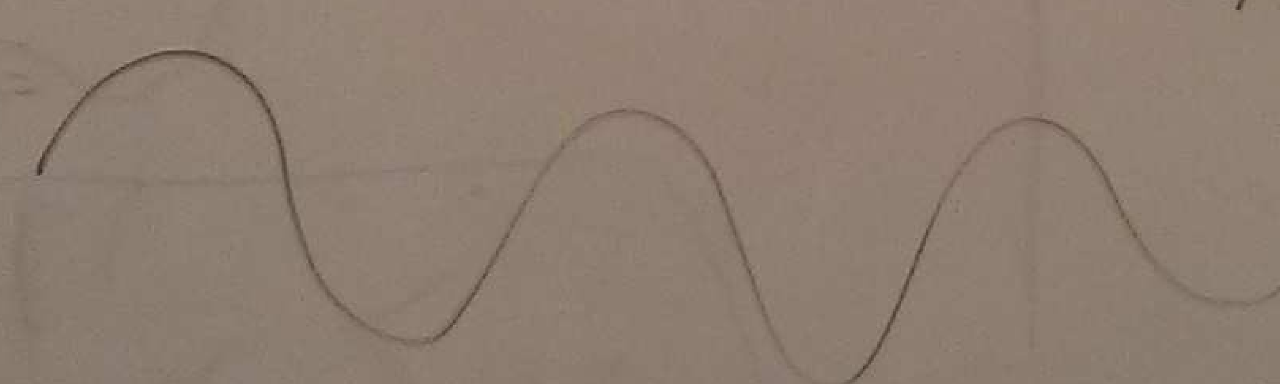
RBF: $x_i \mapsto \exp\left(\frac{-(x_i - x_j)^2}{2\sigma^2}\right) \rightarrow$ go thro

(if e

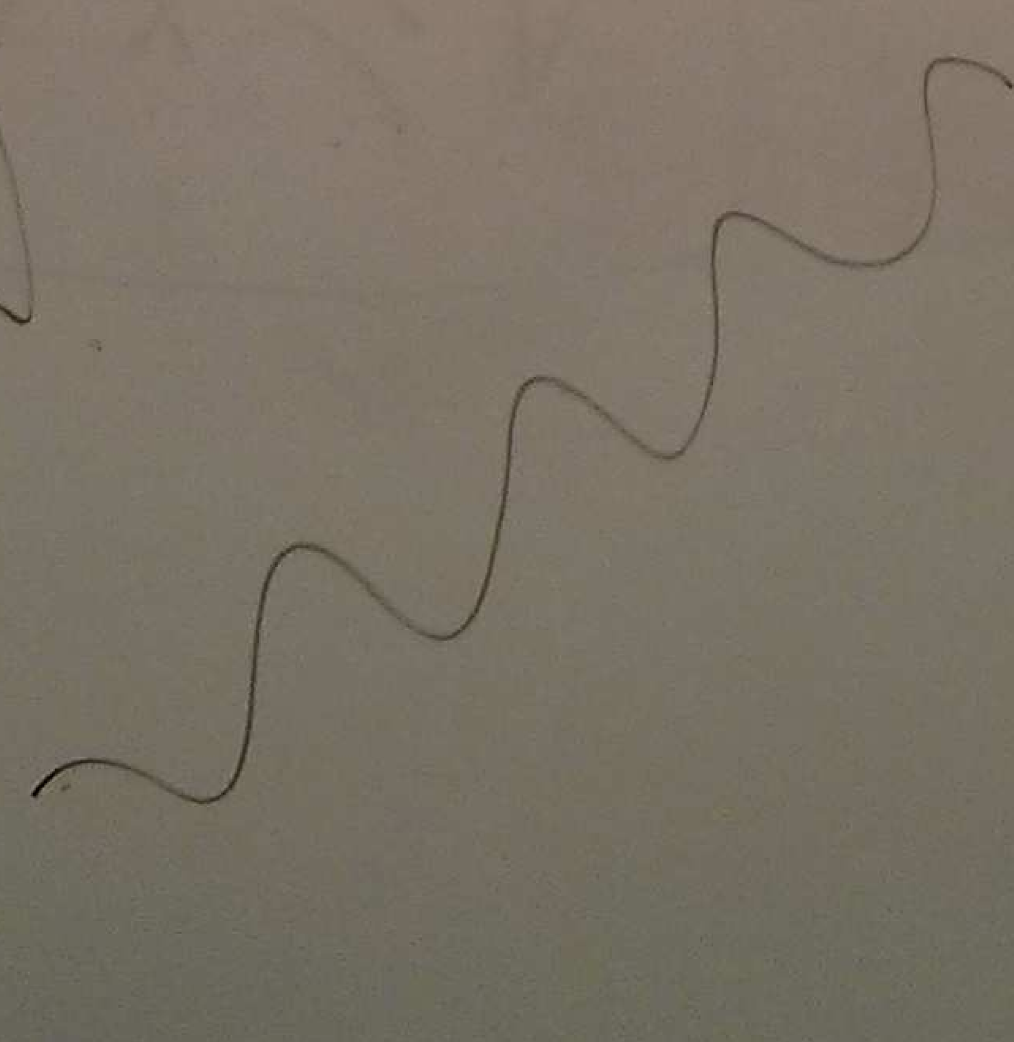
"radial basis function"

"universal approximator"

sin(x)



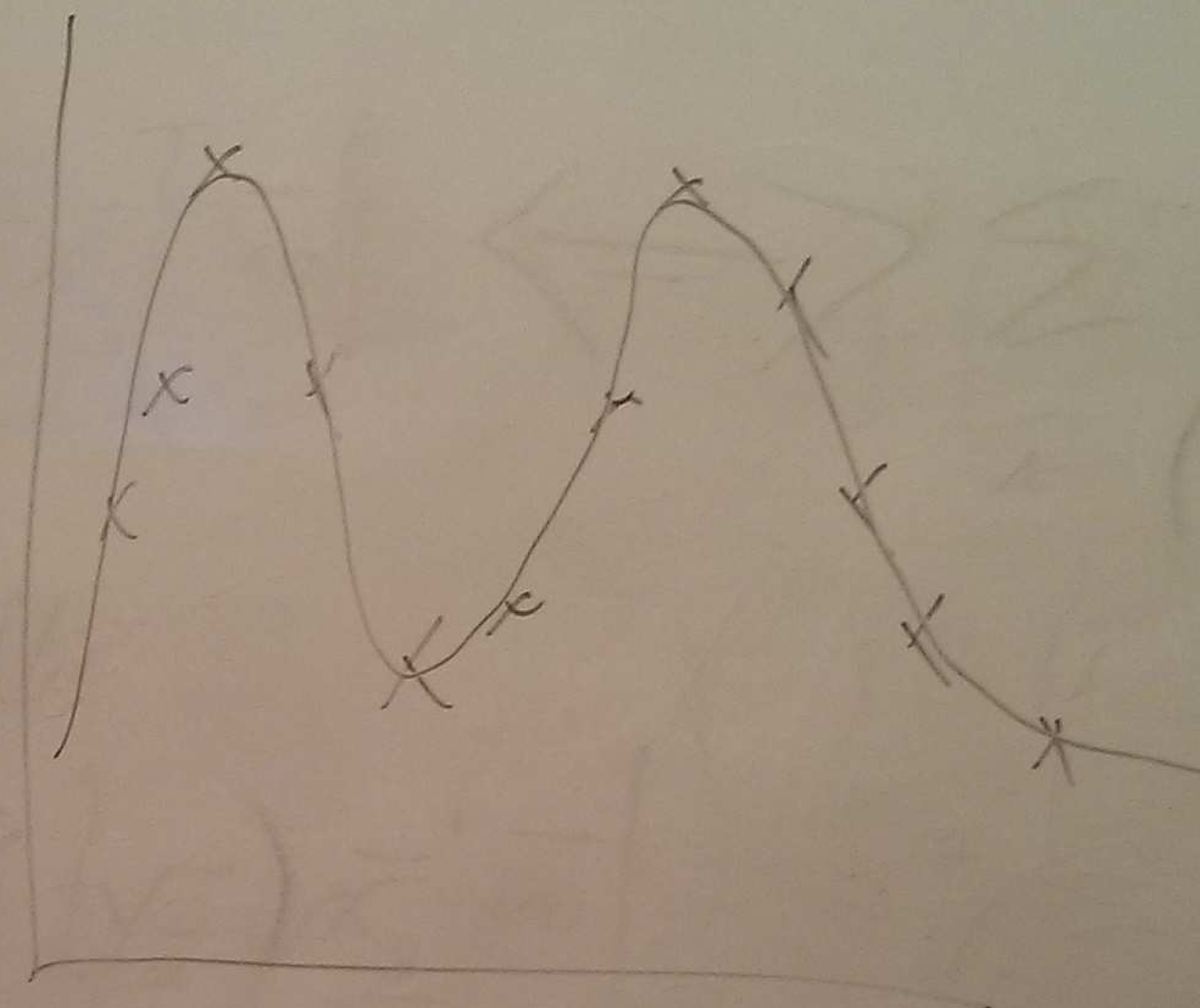
$[x \quad \sin(x)]$



$^{2} \quad x^{3} \; - - - - \; x^{m}]$

$$\frac{- x_j)^2}{2\sigma^2}\Big)$$

"approximator"



→ go through every data point (if error=0) ssion

egression

on

Model Selection

KNN: parameter 'k'

Naive Bayes: "add one"
$\downarrow$
α

Poly basis: parameter 'm'

Which features $\vec{x}_i^j$ to include

Cross Validation

for $k \in \{1, 2, \ldots, K\}$

$$X = \begin{bmatrix} X_{train} \\ X_{validation} \end{bmatrix}$$

- train w/ 'k' on Xtrain
- test on $X_{validation}$

$$X = \begin{bmatrix} X_{validation} \\ X_{train} \end{bmatrix}$$

Akaike information criterion

Bayesian information criterion

put penalty on "degrees of freedom"

information criterion

information criterion

nalty on ees of reedom"

10-fold CV

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ \vdots \\ X_{10} \end{bmatrix}$$

- train on $X_1, \ldots, X_9$
- test on X_{10}

Ridge Regression

$$\sum_{i=1}^{d} w_i^2 = W^T W = \|w\|^2$$

$$\underset{\vec{w} \in \mathbb{R}^d}{\operatorname{argmin}} \frac{1}{2}\sum_{i=1}^{N}(y_i - \vec{w}^T \vec{x}_i)^2 + \frac{\lambda}{2}\sum_{i=1}^{d} w_i^2$$

Why?

- encourage w to be
- smoother