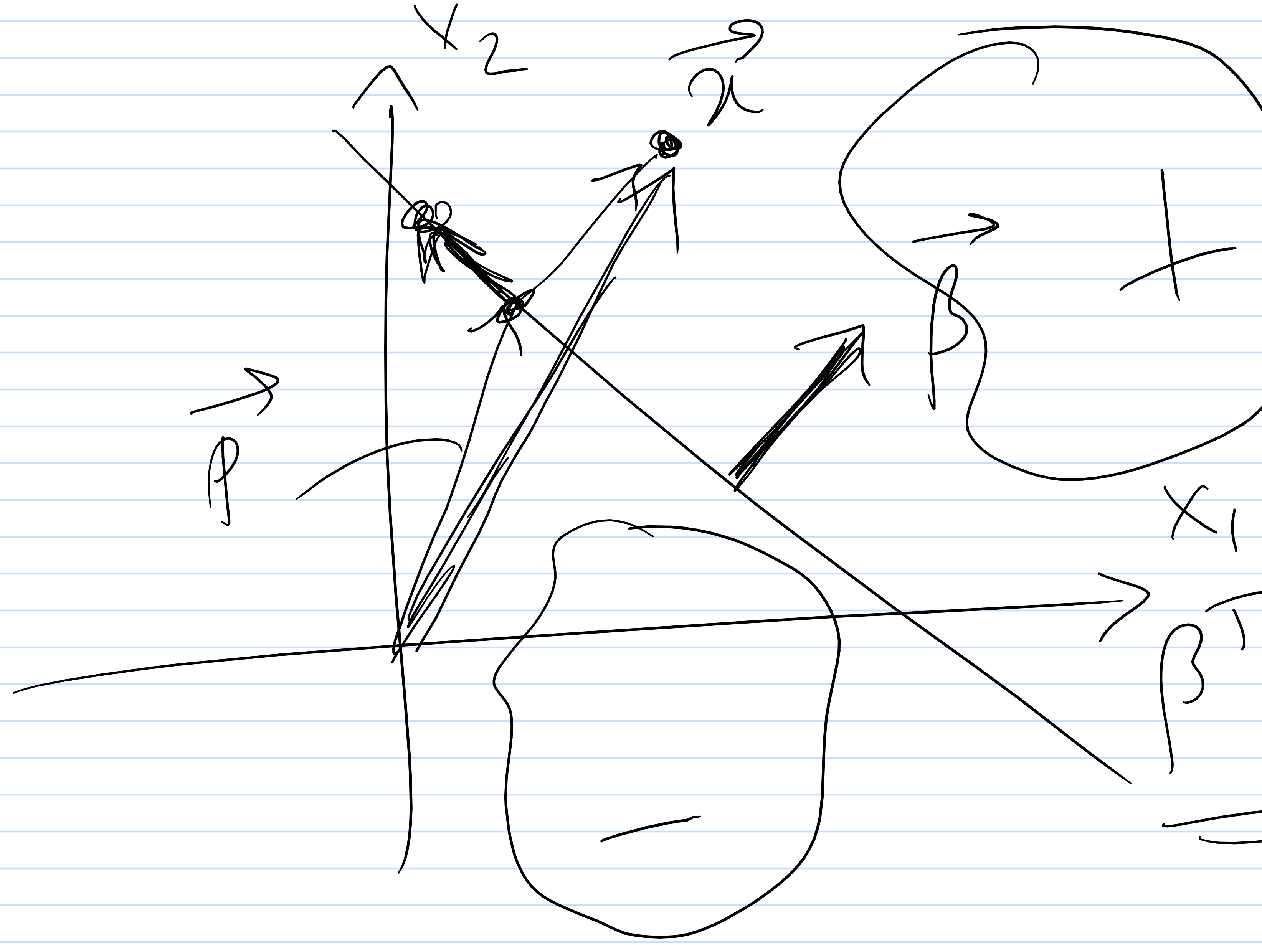
1. Revisit SVM Computation.
2. Extension to nonlinear modeling
3. Unsupervised learning

Problem :

(P) Primal problem

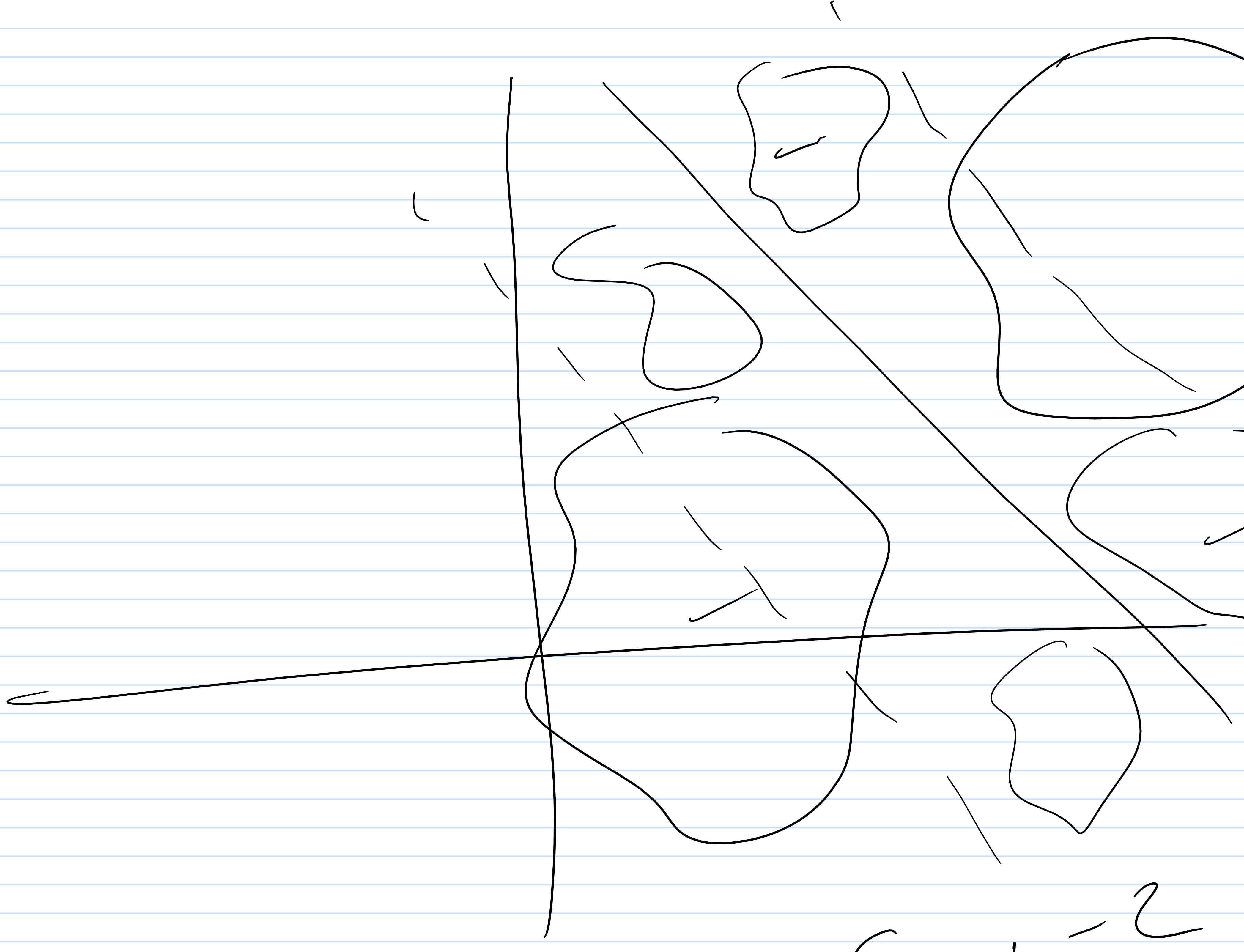
$$\begin{cases} \min_{\beta, \beta_0, \xi_i} \frac{1}{2}\|\beta\|_2^2 + \gamma C \sum_{i=1}^{N} \xi_i = p^* \\ y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i \quad i=1 \dots N \\ \xi_i \geq 0 \quad i=1, \dots N \end{cases}$$

$\vec{p}_1 - \vec{p}_2$



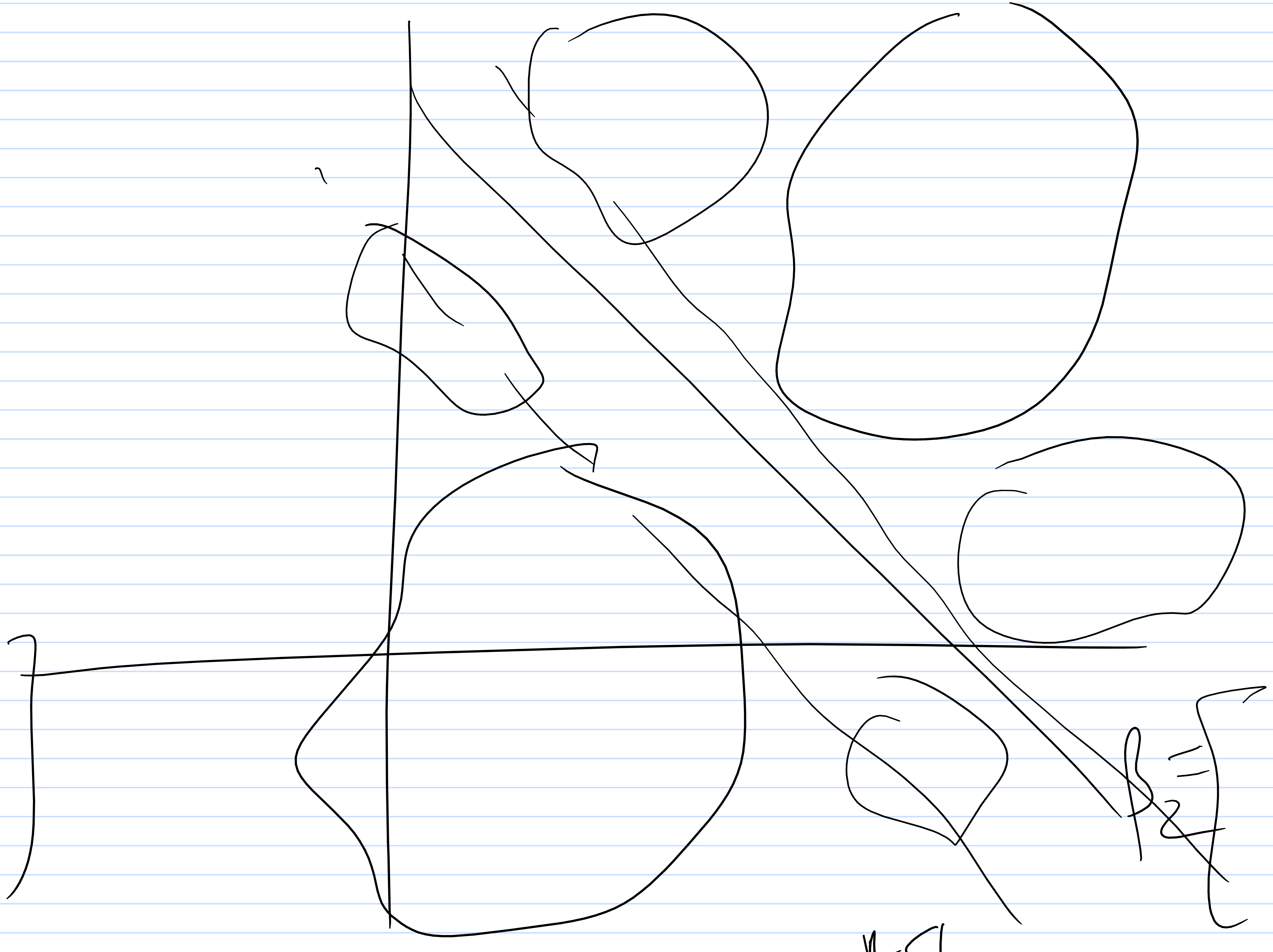
$\underline{\beta^T p + \beta_0 = 0}$

C ? $C = 10^{-2}$ vs $C = 10^{4}$



$\beta_1 = [\ \ ]$

$C = 10^{-2}$



$\beta_2 = [\ \ ]$

$C = 10^{4}$

Lagrange relaxation / Lagrangian

$$\min_{r_i} \sum_{i=1}^{K} q_i \log \frac{r_i}{q_i} = KL(r\|q) = KL(\underline{r}, \underline{q}) \geq 0$$

$$\text{st} \quad \sum_{i=1}^{K} r_i = 1.$$

$$\rightarrow \quad r_i \geq 0 \quad i=1,\dots K$$

Lagrangian : $\underbrace{\sum_{i=1}^{K} q_i \log \frac{r_i}{q_i}}\ + \lambda \underbrace{\left(\sum_{i=1}^{K} r_i - 1\right)}$

$$\frac{d}{dr_i}\left(\quad\right) = \underbrace{\frac{q_1}{r_1}}\ + \lambda = 0 \Rightarrow r_i = \frac{-q_i}{\lambda} \quad k$$

$$\frac{d}{d\lambda}\left(\quad\right) = \sum r_i - 1 = 0 \Rightarrow \sum r_i = 1 \quad 1$$

Solution ? $\boxed{r_i = q_i}$ ☆

$$\sum_i -\frac{q_i}{\lambda} = 1$$

$$1 = \sum q_i = -\lambda \quad \boxed{\lambda = -1} \ ☆$$

Primal

$$\mathcal{L}(\beta, \beta_0, \xi_i, \alpha_i, \mu_i) = \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left(y_i(x_i^T\beta+\beta_0) - 1 + \xi_i\right)$$

$$- \sum_{i=1}^{N}\mu_i(\xi_i - 0)$$

2N

$\boxed{LHS \geq RHS}$

$$\frac{\partial\mathcal{L}}{\partial\beta}, \quad \frac{\partial\mathcal{L}}{\xi_i}$$

$$\frac{\partial\mathcal{L}}{\beta_0}, \quad \frac{\partial\mathcal{L}}{\alpha_i} \quad \frac{\partial\mathcal{L}}{\mu_i}$$

$\alpha_i \geq 0$

$\mu \geq 0$

$\alpha_i = C - \mu_i$

$$\frac{\partial L}{\partial \beta} = \beta - \sum_{i=1}^{M} \alpha_i y_i x_i = 0 \qquad \frac{d}{d\beta}\left(\frac{1}{2}\beta^T \beta\right) = \beta$$

$$\beta = \sum_{i=1}^{M} \alpha_i y_i x_i \quad (1)$$

$$\sum \alpha_i y_i = 0 \quad (2)$$

$$\alpha_i = C - \mu_i \quad (3)$$

$$y_i (x_i^T \beta + \beta_0) \geq 1 - \xi_i \quad (4)$$

$$\xi_i \geq 0 \quad (5)$$

Substitute what we inferred into $\mathcal{L}$

$$\mathcal{L}' = \sum_{i=1}^{N} \underline{\alpha}_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

2nd problem

Dual problem:

$$\max_{\alpha_i} \mathcal{L}'(\underline{\alpha}) = d^*$$

st

$$0 \leq \alpha_i \leq C \quad i = 1 \ldots N$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

fact 1 : $\underset{\beta^*, \beta_0^*, \xi_i^*}{P^*} \geq \underset{\alpha_i^*, \mu_i^*}{d^*}$: weak duality

fact 2 : $P^* = d^*$: strong duality

min $f(\beta)$

st $g(\beta) \geq 0$

f convex

$g(\beta) \geq 0$

g concave

fact 3 : If strong duality

then $\alpha_i^*\left(y_i(x_i^T\beta^* + \beta_0^*) - 1 + \xi_i\right) = 0$

$(C - \alpha_i^*)\xi_i = 0$

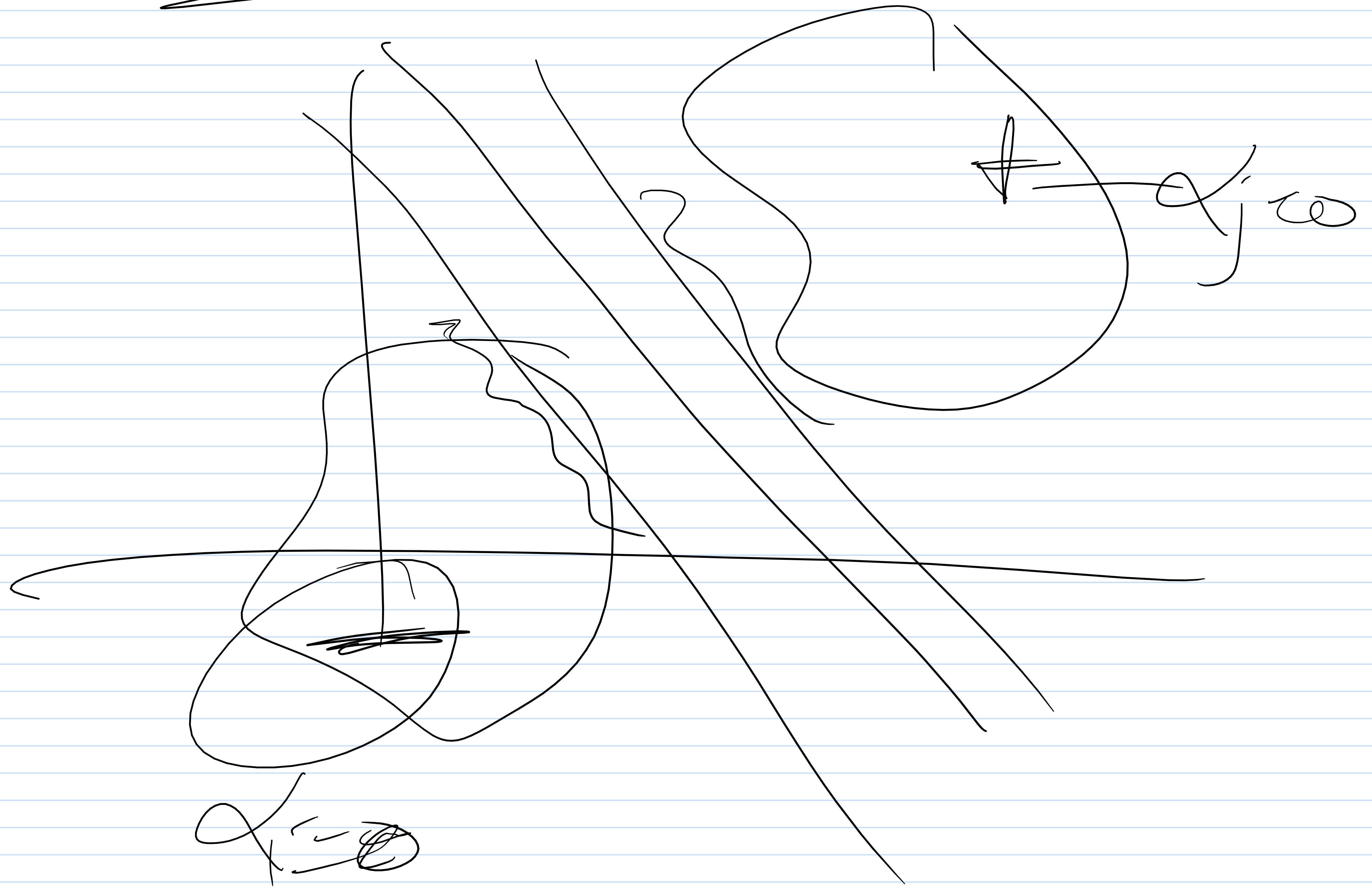
Complementarity Conditions

$$\alpha_i \left(y_i (x_i^T \beta + \beta_0) - 1 + \xi_i \right) = 0$$

i : α_i is non zero.

$$y_i (x_i^T \beta + \beta_0) = 1 - \xi_i$$

$$\underset{p\times 1}{\beta} = \sum_{i=1}^{N} \alpha_i y_i \underset{p\times 1}{x_i}$$



Getting a nonlinear model.

β_0

$$\beta = \underbrace{\sum_{i=1}^{N} \alpha_i y_i x_i}$$

$$f(\underline{x}) = \beta^T x = x^T \beta$$

$$= \underline{\sum_{i=1}^{N} \alpha_i y_i \, \underline{x^T x_i}}$$

$$\boxed{\underline{\sum_{i=1}^{N} \alpha_i y_i K(\underline{x}^{test}, x_i)}}$$

$$x^T x_i$$

$$\downarrow$$

$$\underbrace{\underline{\phi(x)}^T \underline{\phi(x_i)}} = \underline{K}(x, x_i)$$

Kernel function.

Kernel trick.

ϕ. $K(x_i, x_j)$

eg: $e^{-\gamma \|x_i - x_j\|_2^2}$

eg1: $(x_i^T x_j + 1)^d$ $d = 2$

$= K(x_i, x_j)$

$=$ expand. $(x_i^T x_j)^2 + 1 + 2 x_i^T x_j$

$\longleftarrow \phi(x_i)^T \phi(x_j)$

$$\begin{pmatrix} 1 \\ \sqrt{2} x_{i1} \\ \sqrt{2} x_{i2} \\ \vdots \\ \sqrt{2} x_{ip} \\ x_{i1}^2 \\ \vdots \end{pmatrix}$$

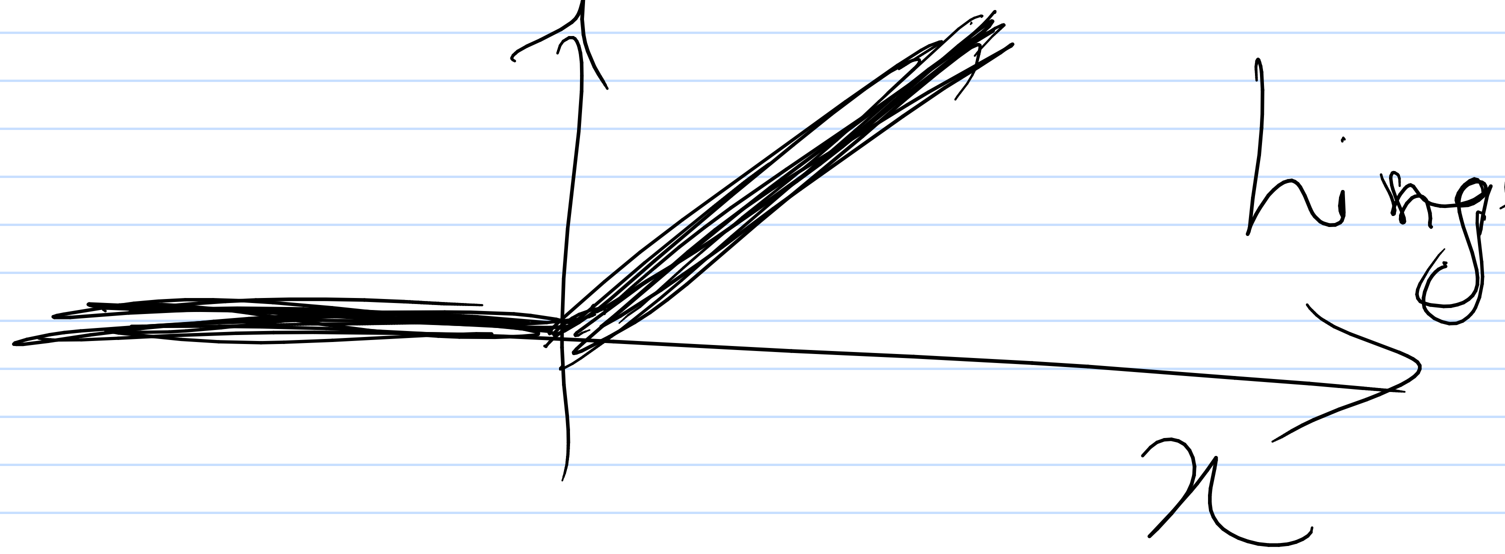
$$\min_{\beta,\beta_0,} \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^{N} \max\left(0, 1 - y_i\left(x_i^T\beta+\beta_0\right)\right)$$

$$y_i\left(x_i^T\beta+\beta_0\right) \geq 1-\xi_i \Longrightarrow \xi_i \geq 1 - y_i\left(x_i^T\beta+\beta_0\right)$$

$$\xi_i \geq 0 \longrightarrow \xi_i \geq 0$$

$$\Rightarrow \xi_i \geq \max\left(0, 1-y_i\left(x_i^T\beta+\beta_0\right)\right)$$

$\max(0, x)$

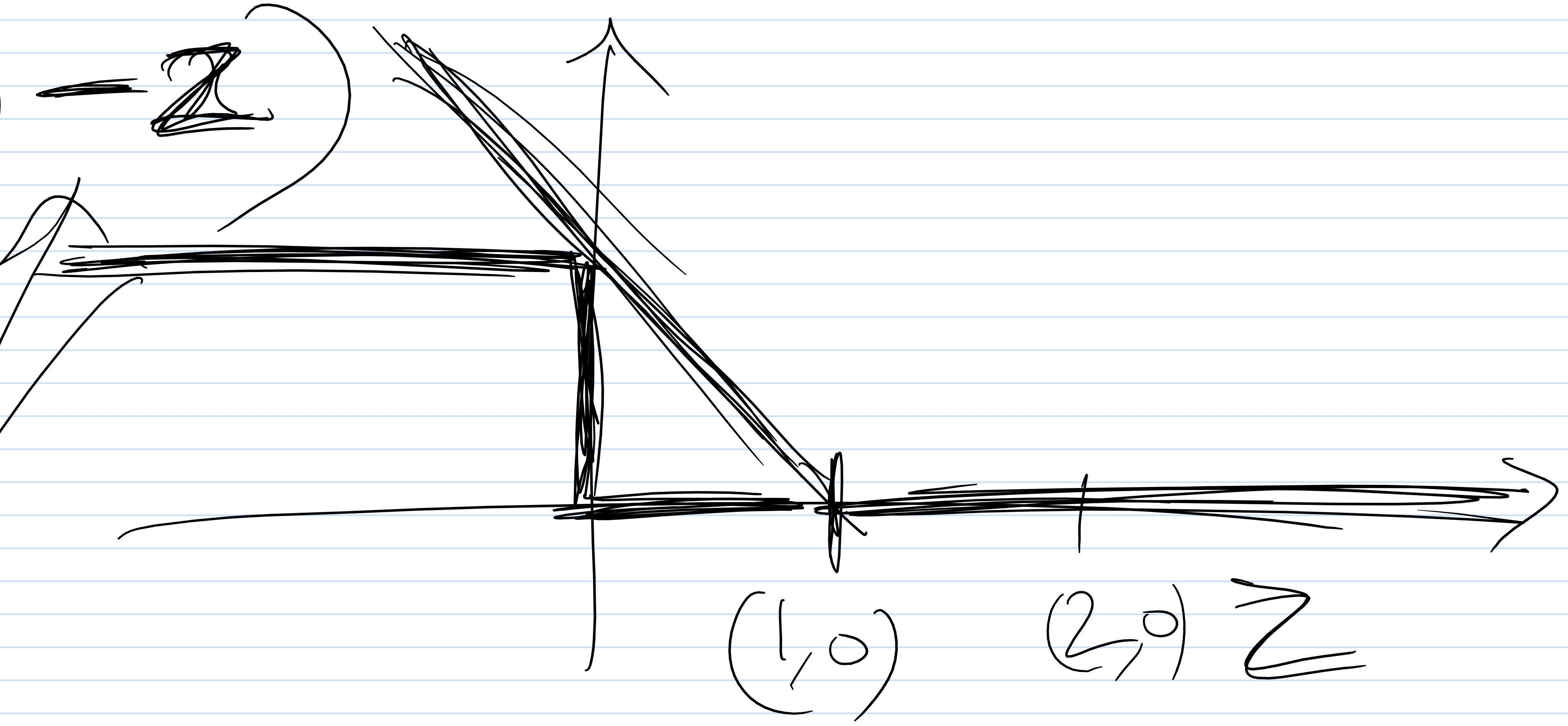


hinge function

ReLU

$\max(0, 1-z)$

convex



0-1 loss function

$\mathbb{1}[z < 0]$

is not

$y \in \{-1, 1\}$

$\text{sign}(f(x)) \in \{-1, 1\}$

$y f(x) > 0$

$y f(x) < 0$

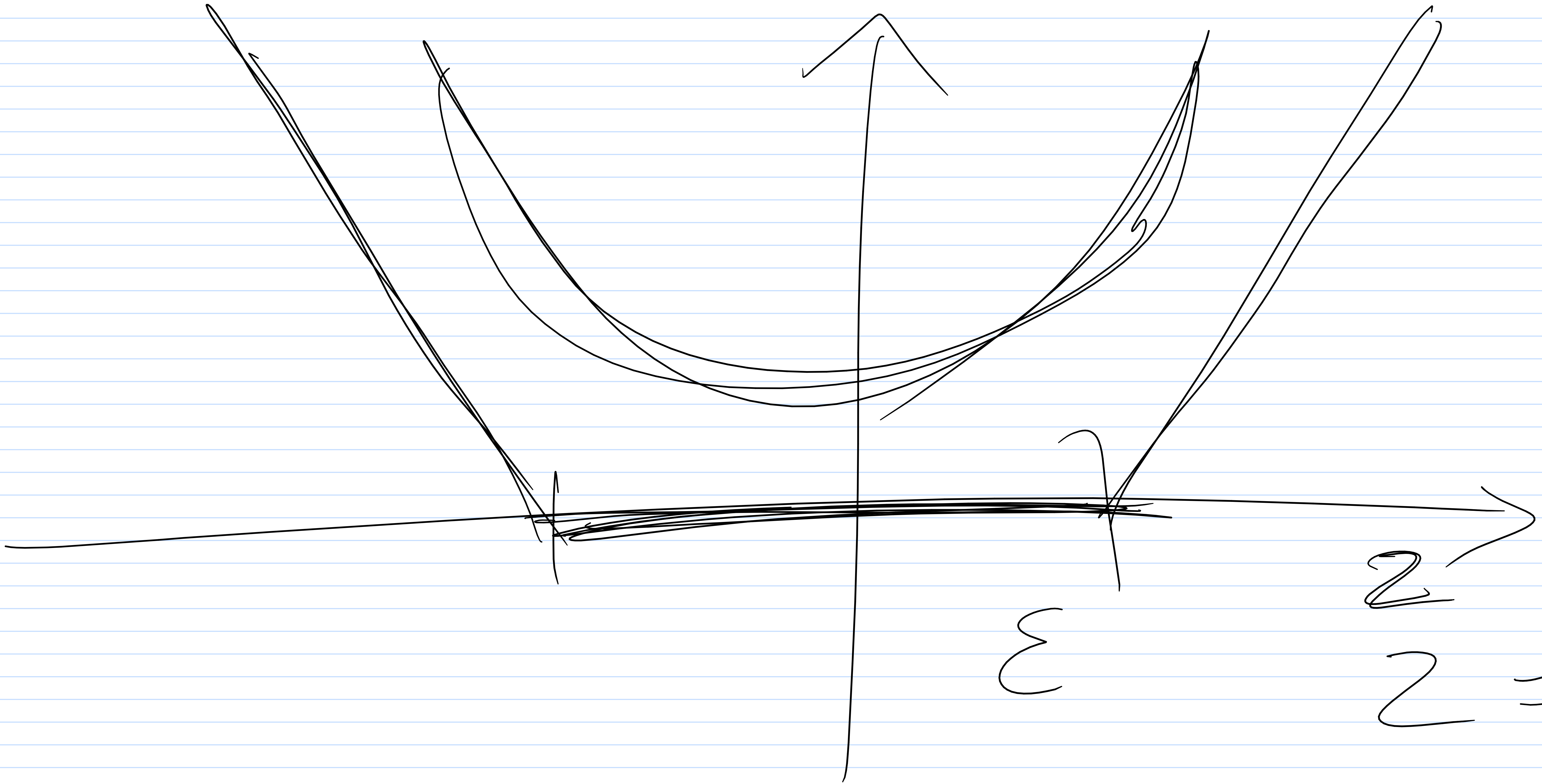
$$\min_{\beta, \beta_0} \|\beta\|_2^2 + C \cdot \text{hinge loss (data)}$$

$\downarrow$

Support vector regression

ε - insensitive loss

$$l(\cancel{f}f(x), y) = \begin{cases} |y - f(x)| - \varepsilon & \text{if } |y - f(x)| > \varepsilon \\ 0 & \text{otherwise} \end{cases}$$



$$z = y - f(x)$$

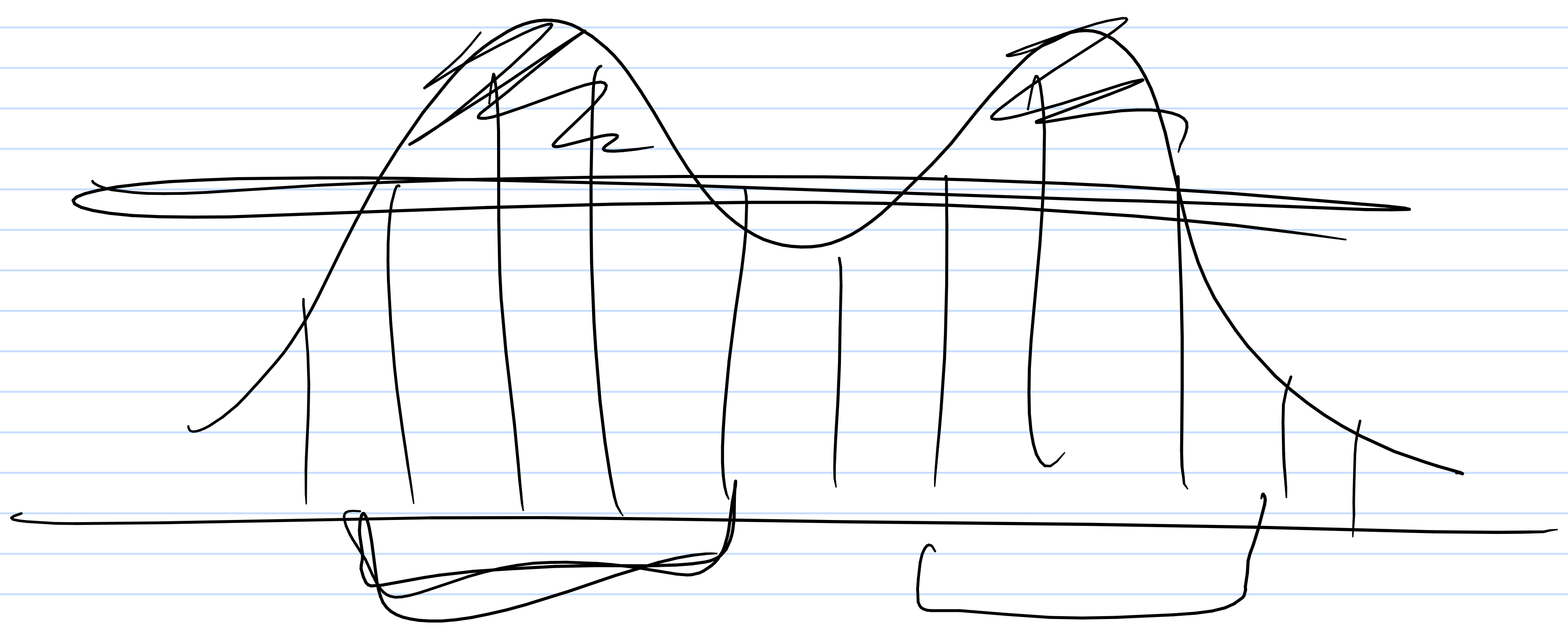
Unsupervised learning.

y_i x_i

$\hat{y}_i$

eg MLE, EM

GMM



Goal : exploratory

: No objective measure of success

understand $P(X)$

Principal Components;

clustering : k-means, spectral clustering

association rules : frequent itemsets

GMM

data EM

$$\mu_1 \quad \mu_2$$
$$\sigma_1 \quad \sigma_2$$
$$\pi$$

x $\quad P(X = x \mid \text{1st component})$

$P(X = x \mid \text{2nd component})$

AR

$$\overbrace{X_1 \ldots\ldots X_p}^{X}$$

binary dataset $X_1 \ldots\ldots X_p$ $\quad X_i \in \{0,1\}$

Goal : find joint values of subsets of X that appeared most frequently in the dataset.

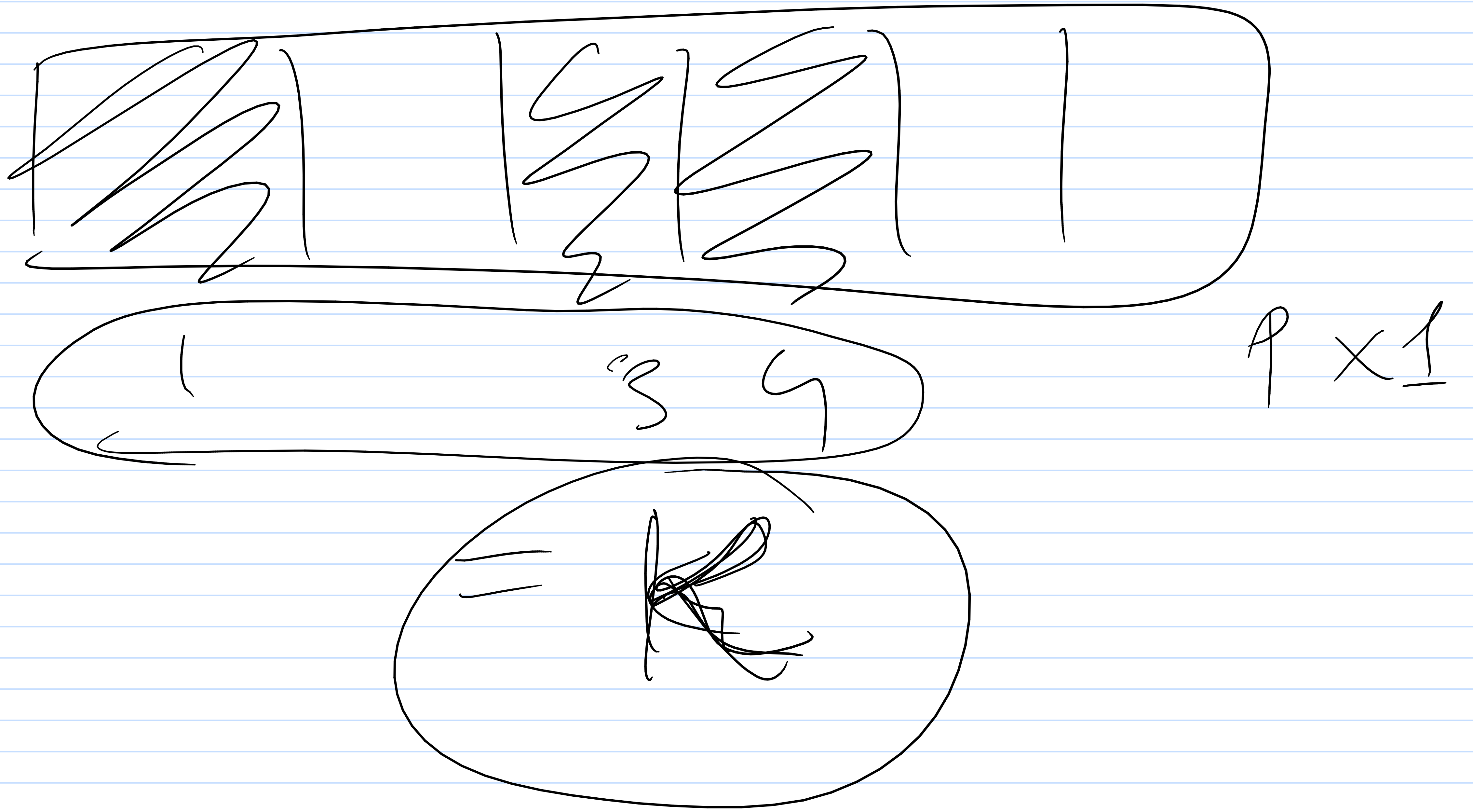
$S_i \subseteq \{0,1\}$ $\quad \{0\}, \{1\}, \{0,1\}, \emptyset$

find $S_1 \ldots\ldots S_p$ st

$$\Pr\left(\bigcap_{i=1}^{p} \{X_i \in S_i\}\right) \text{ is high}$$

$$\Pr\left(\prod_{k \in K} X_k = 1\right)$$

high



estimate from data.

support of K

$$\frac{1}{N}\sum_{i=1}^{N}\prod_{k \in K} x_{ik}$$

(A) $\{K_j : \text{support}(K_j) > 0.1\}$: frequent itemsets

$\downarrow$

Rules (AR)

$K_j = \{1 \quad 3 \quad 4\}$

$\boxed{\{1\} \quad \{3,4\}}$

$\{1,3\} \quad \{4\}$

$\{1\ 4\} \quad \{3\}$

$\{3\ 4\} \quad \{1\}$

$\boxed{\text{If } \{1\} \text{ then } \{3,4\}}$

pick rules with high confidence.

K_1

$$K_1 \rightarrow \{ p \Rightarrow q \} \quad \frac{\text{Support}(p \cup q)}{\text{Support}(p)}$$

$\{1, 3, 4\}$

$$\approx P(X_3 = 1, X_4 = 1 \mid X_1 = 1)$$

How to get $\{K_j : \text{support}(K_j) > 0.1\}$

Apriori

Core idea: if $K_1 \subseteq K_2$ then $\text{support}(K_1) \geq \text{support}(K_2)$

Work in rounds

1st round: Single-item sets

if > 0.1 keep

2nd round : form pairs from the selected single item sets from previous

$\text{Support}(K_j) > 0.1$

mth round : form m sized sets from previous $(m-1)$ sized sets.

$\{1\ 5\ 9\ 11\} \longleftarrow \left\{ \begin{array}{c} 1\ 5\ 9 \\ 1\ 5\ 11 \\ 1\ 9\ 11 \\ 5\ 9\ 11 \\ \vdots \end{array} \right\}$