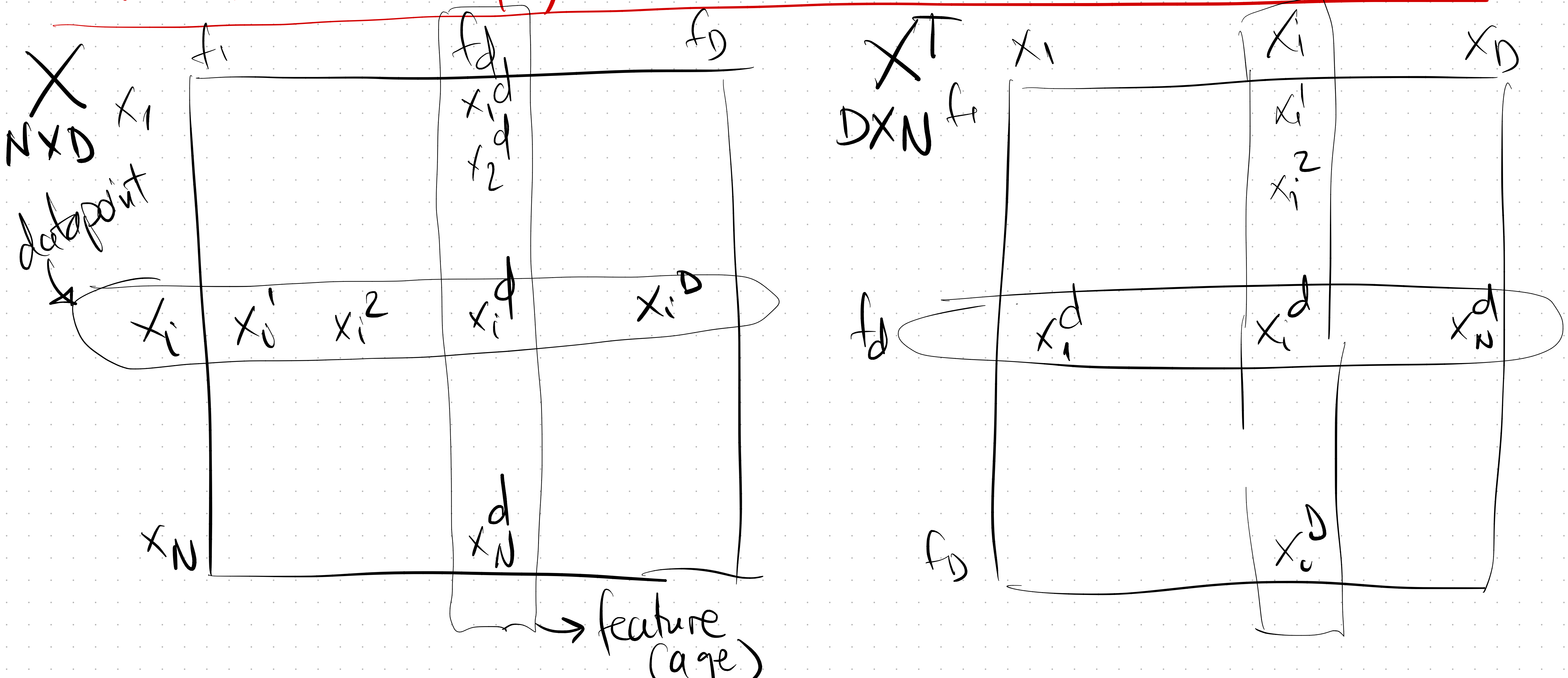
Lecture 5/16

- X^TX matrix $\Rightarrow (X^TX+\lambda I)^{-1}$ . . .
- Linear Separability HW1 pb5 / vs Convex Hulls
- HW1 pb4: Entropy, Cond Entropy, Mutual Information



X [0-mean columns] $\Rightarrow$ X is $\boxed{\text{centered}}$

$$X \quad \begin{array}{ccc} x^1 & x^d & x^D \\ \hline & & \\ & & \\ \hline \end{array}$$

$$\mu \quad \begin{array}{ccc} \mu^1 & \mu^d & \mu^D \end{array}$$

D-dim

$\mu = \text{mean}(X)$

$\mu^d = \text{mean}(X\_\text{column}\,d)$

$$\xrightarrow{\text{CENTER}} \quad \begin{bmatrix} x_1^1 - \mu^1 & x_1^2 - \mu^2 & \cdots & x_1^D - \mu^D \\ x_2^1 - \mu^1 & x_2^2 - \mu^2 & \cdots & x_2^D - \mu^D \\ \vdots & & & \\ x_N^1 - \mu^1 & x_N^2 - \mu^2 & \cdots & x_N^D - \mu^D \end{bmatrix}$$

$\mu_{(X\_\text{CENTERED})} = 0$ on each col

X has 0-mean columns

X centered. look at $X^T X$

$$
\begin{array}{c|cccc|}
 & & & \text{feat } h & \\
 & \underbrace{\sum_{i=1}^{N} (x_i^1 \cdot x_i^1)}_{\text{var}(x^1)} & & & \sum_{i=1}^{N} x_i^1 \cdot x_i^D \\
 & & \text{var}(x^2) & & \\
\text{feat } d & & & \sum_{i=1}^{N} x_i^d \cdot x_i^h & \\
h & & & \underbrace{\sum_{i=1}^{N} x_i^h \cdot x_i^h}_{\text{VAR}(x^h)} & \\
D & \sum_{i=1}^{N} x_i^D \cdot x_i^1 & & & \underbrace{\sum_{i=1}^{N} x_i^D \cdot x_i^D}_{\text{VAR}(x^D)} \\
\hline
 & & D & &
\end{array}
=
\begin{array}{c|ccccc|}
 & & & f & & \\
 & \langle x^1 \cdot x^1 \rangle & \langle x^1 \cdot x^2 \rangle & \cdots & & \langle x^1 \cdot x^D \rangle \\
d & & & \langle x^d \cdot x^f \rangle & & \\
 & \langle x^D \cdot x^1 \rangle & & & & \langle x^D \cdot x^D \rangle
\end{array}
$$

so in coll (d, f)

we have $x^d \cdot x^f =$

$\langle \underline{\text{feature\_d}} \times \underline{\text{feature\_f}} \rangle$

$=$ Linear Correlation $\left(\begin{array}{c} \text{feat } d \\ \text{feat } f \end{array} \right)$

if $\mu \neq 0$

$$
\begin{array}{c|c}
 & \qquad\qquad d \qquad\qquad h \\
 & \begin{bmatrix}
\sum_{i=1}^{N}(x_i^1-\mu^1)(x_i^1-\mu^1) & \cdots & \cdots & \sum_{i=1}^{N}(x_i^1-\mu^1)(x_i^D-\mu^D) \\
 & & & \\
\end{bmatrix} \\
d & \qquad\qquad\qquad \boxed{\sum_{i=1}^{N}(x_i^d-\mu^d)(x_i^h-\mu^h)} \\
 & \qquad\qquad\qquad \| \\
h & \qquad \boxed{\sum_{i=1}^{N}(x_i^h-\mu^h)(x_i^d-\mu^d)} \\
 & \begin{bmatrix}
\sum_{i=1}^{N}(x_i^D-\mu^D)(x_i^1-\mu^1) & \cdots & \cdots & \sum_{i=1}^{N}(x_i^D-\mu^D)(x_i^D-\mu^D)
\end{bmatrix}
\end{array}
$$

$$= N \cdot \underset{D\times D}{\mathrm{COVAR}(X)}$$

where cell (d,h) is linear correl of feat_d $\times$ feat_h if X centered $(\mu=0) \Rightarrow X^TX = \boxed{N} \cdot \mathrm{COVAR}(X)$

Very nice-math properties. $\mathrm{COVAR} \simeq X^TX$:

- symmetric $X^TX_{i,j} = X^TX_{j,i}$ = lin_correlation(sim) feat_i, feat_j

- pos semidefinite : $\forall$ vector (D-dim) $Z=(z_1 z_2 \dots z_D)$: $\underline{Z^T(X^TX)Z \geq 0}$

  <u>proof</u>: $Z^T(X^TX)Z = Z^TX^T \cdot XZ = (XZ)^T \cdot XZ \geq 0$

$e_i^T \cdot e_j = 0$

- <u>Spectral decomposition</u> $\begin{cases} e_1, e_2, \dots e_D \rightarrow \text{column eigenvectors} \\ \alpha_1 \quad \alpha_2 \quad \quad \alpha_D \rightarrow \text{corresp eigenval} \end{cases}$

- Eigenvectors Orthonormal $\langle e_i^T e_i \rangle = 1$

$(X^TX)e_j = \alpha_j \cdot e_j \Leftarrow$

$$X^TX = \underbrace{\begin{bmatrix} | & | & & | \\ e_1 & e_2 & \dots & e_D \\ | & | & & | \end{bmatrix}}_{E} \cdot \underbrace{\begin{bmatrix} \alpha_1 & & 0 \\ & \alpha_2 & \\ 0 & & \ddots \\ & & & \alpha_D \end{bmatrix}}_{\mathrm{diag}(\alpha)} \cdot \underbrace{\begin{bmatrix} — & e_1 & — \\ — & e_2 & — \\ & \vdots & \\ — & e_D & — \end{bmatrix}}_{E^T}$$

D columns, D rows

Kernel Matrix

$$\underset{N \times D}{X} \cdot \underset{D \times N}{X^T} = \underset{N \times N}{K}$$

2 datapoints

cell ij: $\boxed{K_{ij} = \langle x_i \cdot x_j \rangle = \text{linear similarity}(x_i, x_j)}$

$$\begin{array}{c|ccccc|}
 & 1 & & j & & N \\
1 & x_1 \cdot x_1 & & x_1 \cdot x_2 & & x_1 \cdot x_N \\
 & & & & & \\
i & x_i \cdot x_1 & \cdots & x_i \cdot x_j & \cdots & x_i \cdot x_N \\
 & & & & & \\
N & x_N \cdot x_1 & & & & x_N \cdot x_N
\end{array}$$

$\|x_1\|^2$ $\|x_2\|^2$ $\|x_N\|^2$

$N \times N$

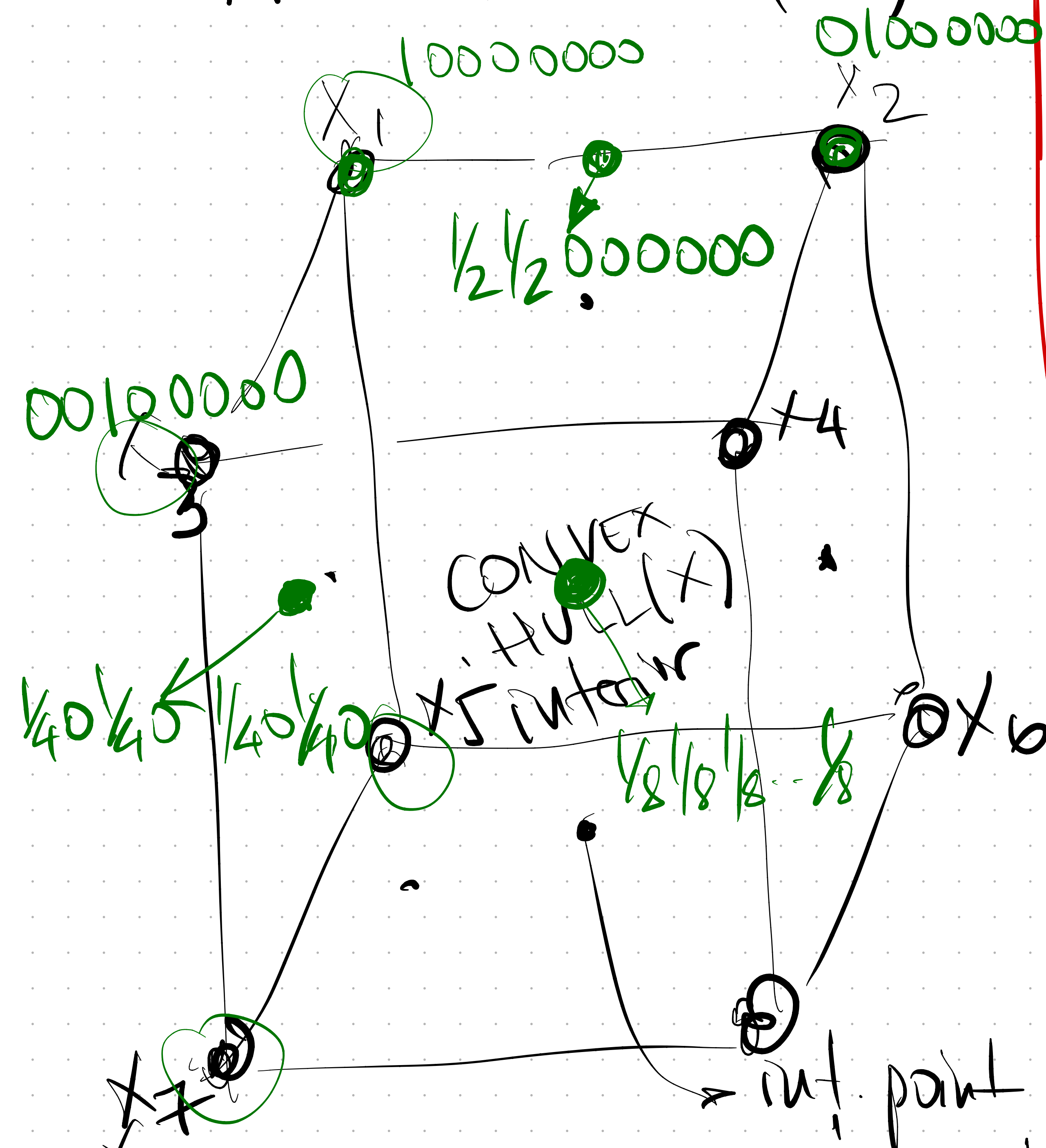
→ pairwise similarity of all datapoints

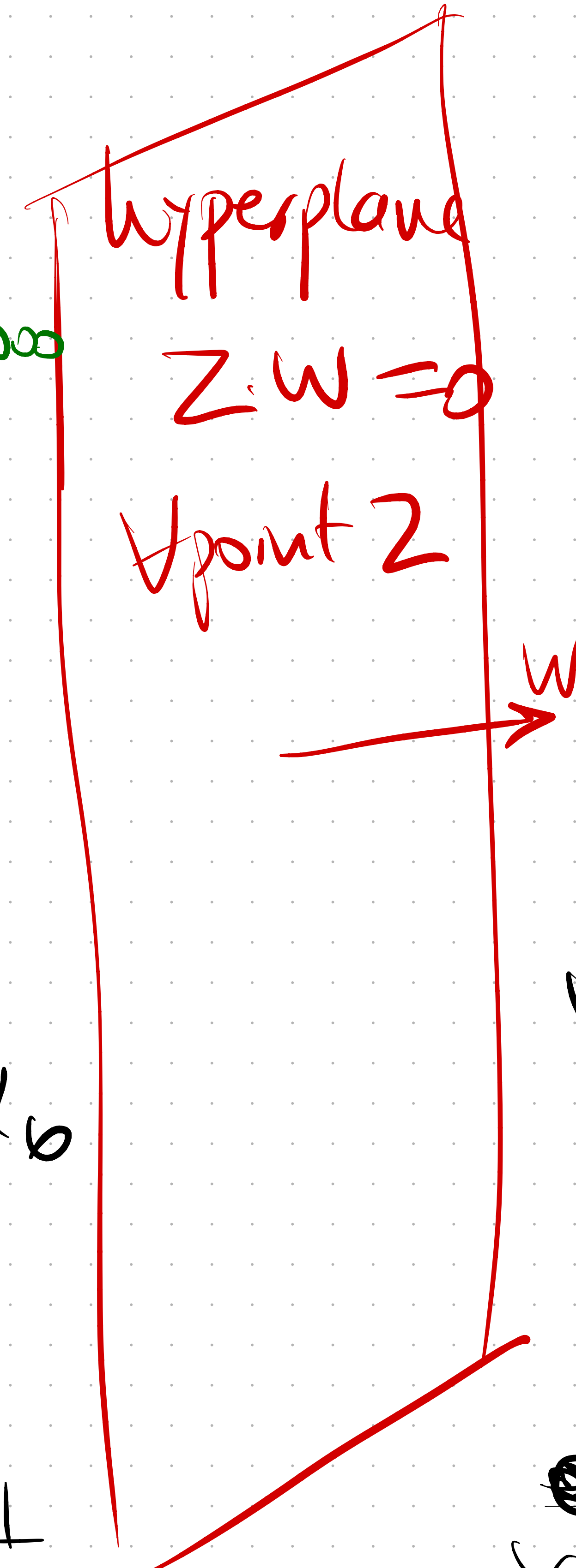
ALGEBRA: $X \cdot X^T$ has same property as $X^T \cdot X$ (matrix · matrix^T)

- symm
- pos def
- eigen spectral decomp

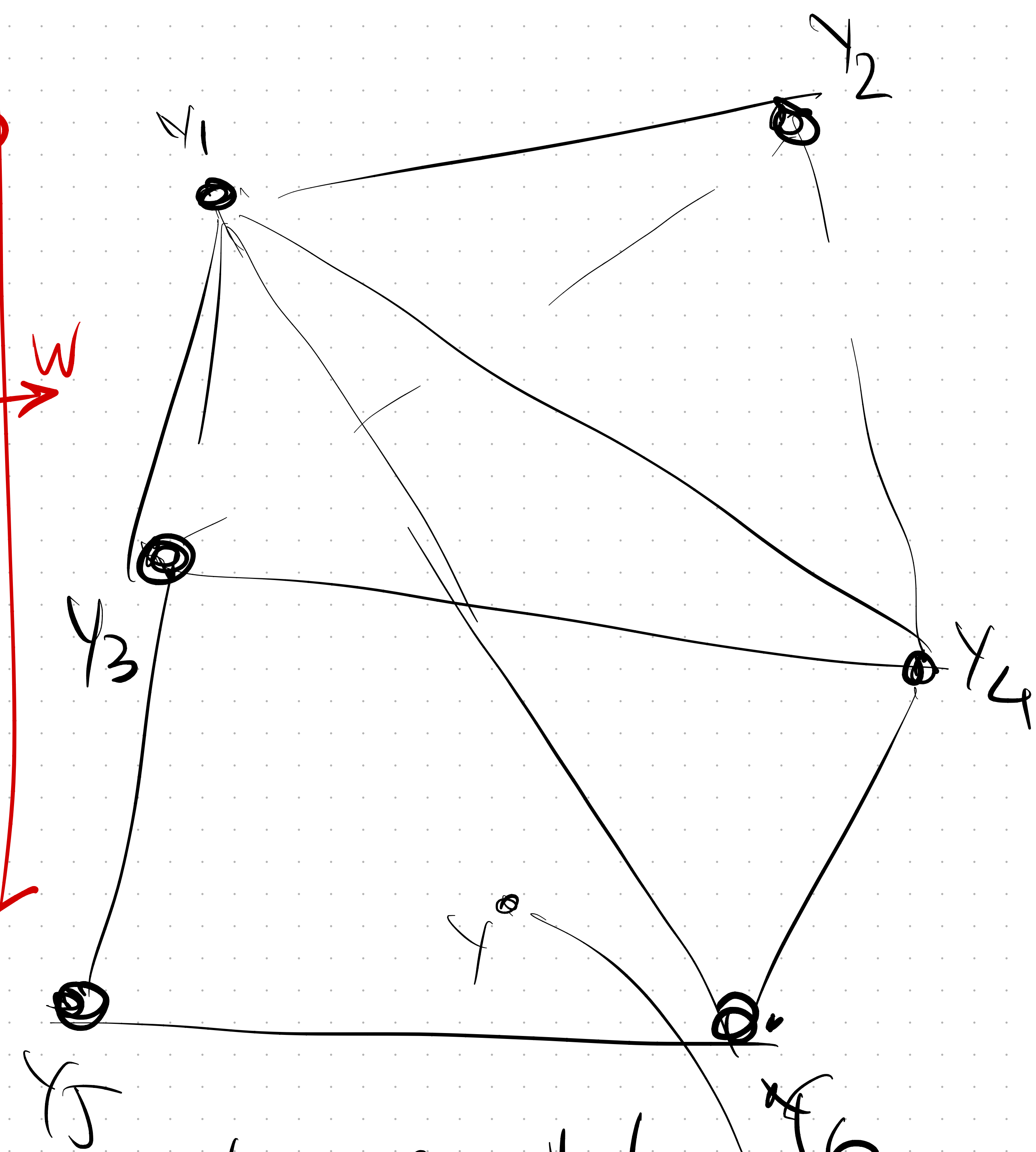
HW1 PB5:

X = set of vectors (3-D)





Y = set of vectors



$x = \sum_{i=1} \alpha_i x_i$

α-dist
$\sum \alpha = 1$
all α possi

convex polyhedra

$CH(X) = \{ \forall \alpha \text{ weights } \sum \alpha_i x_i \}$

convex polyh

$CH(Y) = \{ \forall \alpha \text{ weights } \sum \alpha_i Y_i$

A) either $CH(X)$, $CH(Y)$ donot intersect
or $\{x_i\}$ $\{y_j\}$ are linearly separalle

hyperplane sep: one side X, the other side Y

cannot have both
- lin separation by hyp.
- $CH(X) \cap CH(Y) \neq 0$
  intersection.

B) one of the two must happen

if $CH(X) \cap CH(Y) = \emptyset$ $\Longrightarrow$ there is a separator hyperplane
do not intersect