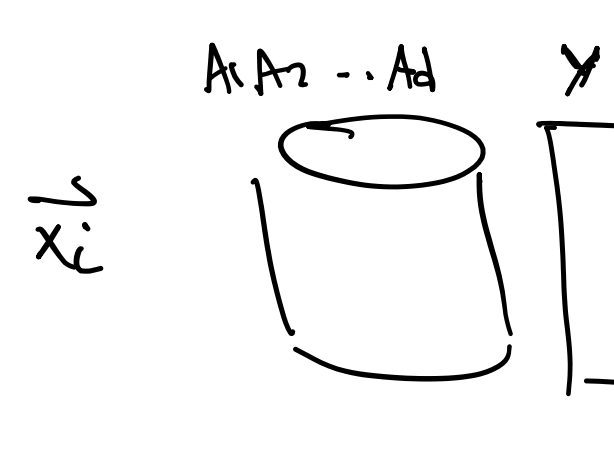
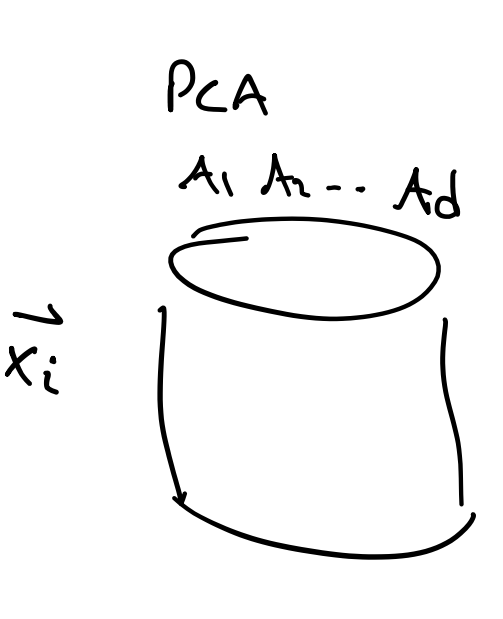
Lecture 14

Monday, October 28, 2024   10:02 AM

Linear Discriminant Analysis   vs   PCA



find a direction to best separate our the classes

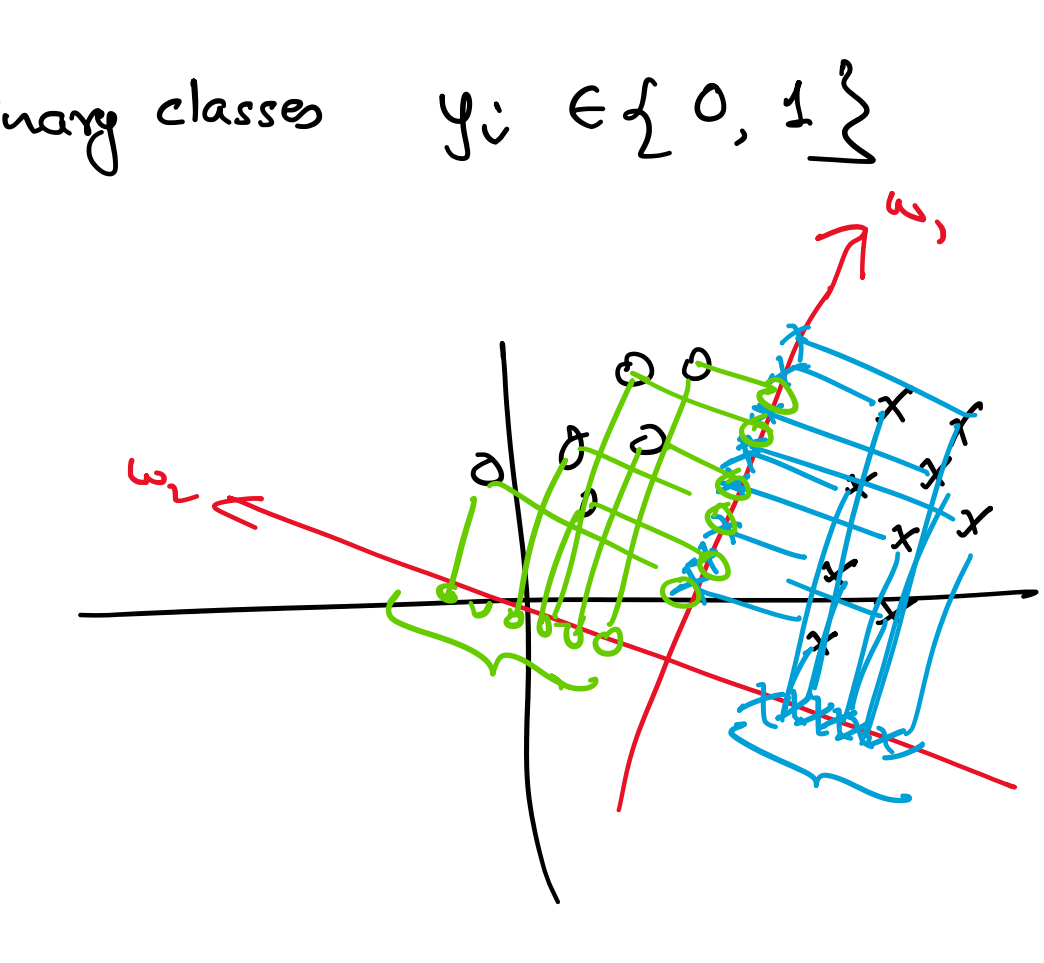


find a direction to best approx the data

minimize MSE

(maximize variance)

binary classes $y_i \in \{0, 1\}$



LDA objective

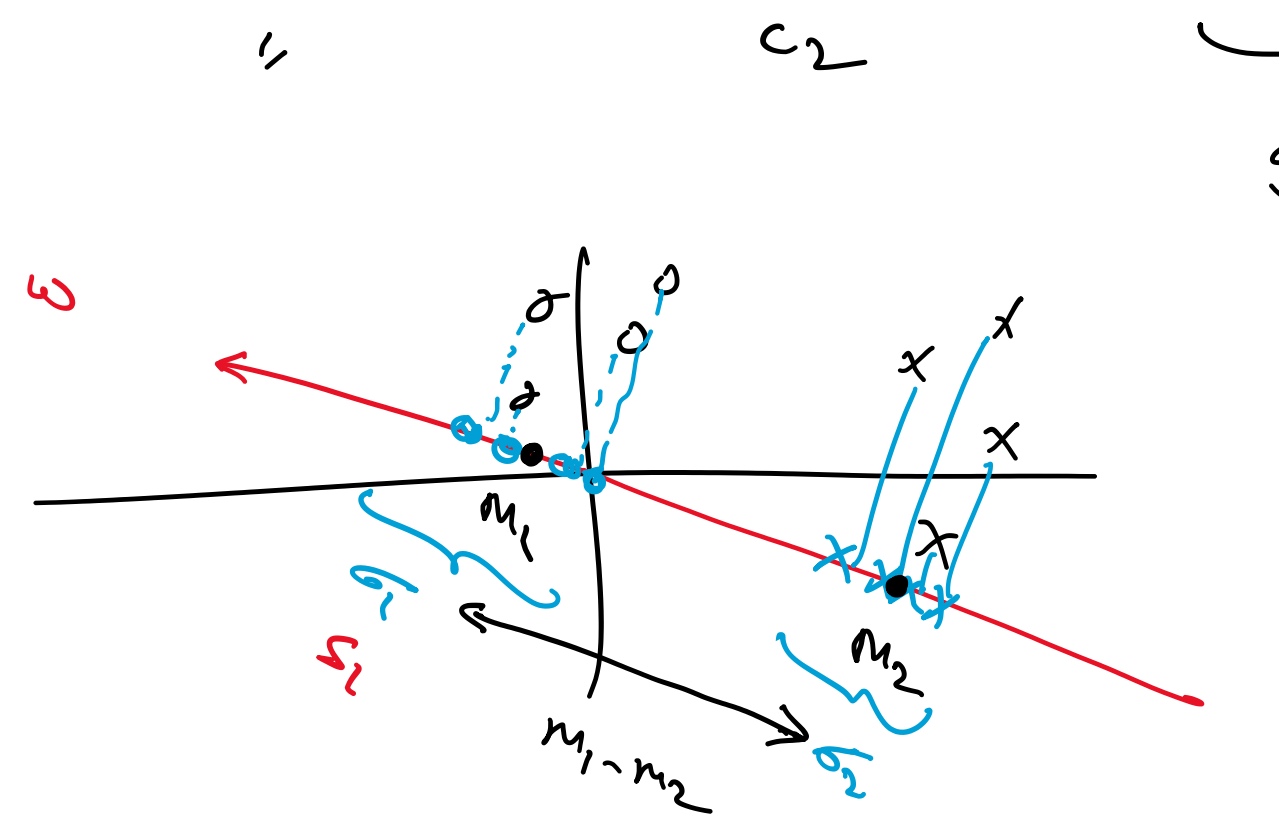
① maximize $(m_1 - m_2)^2$

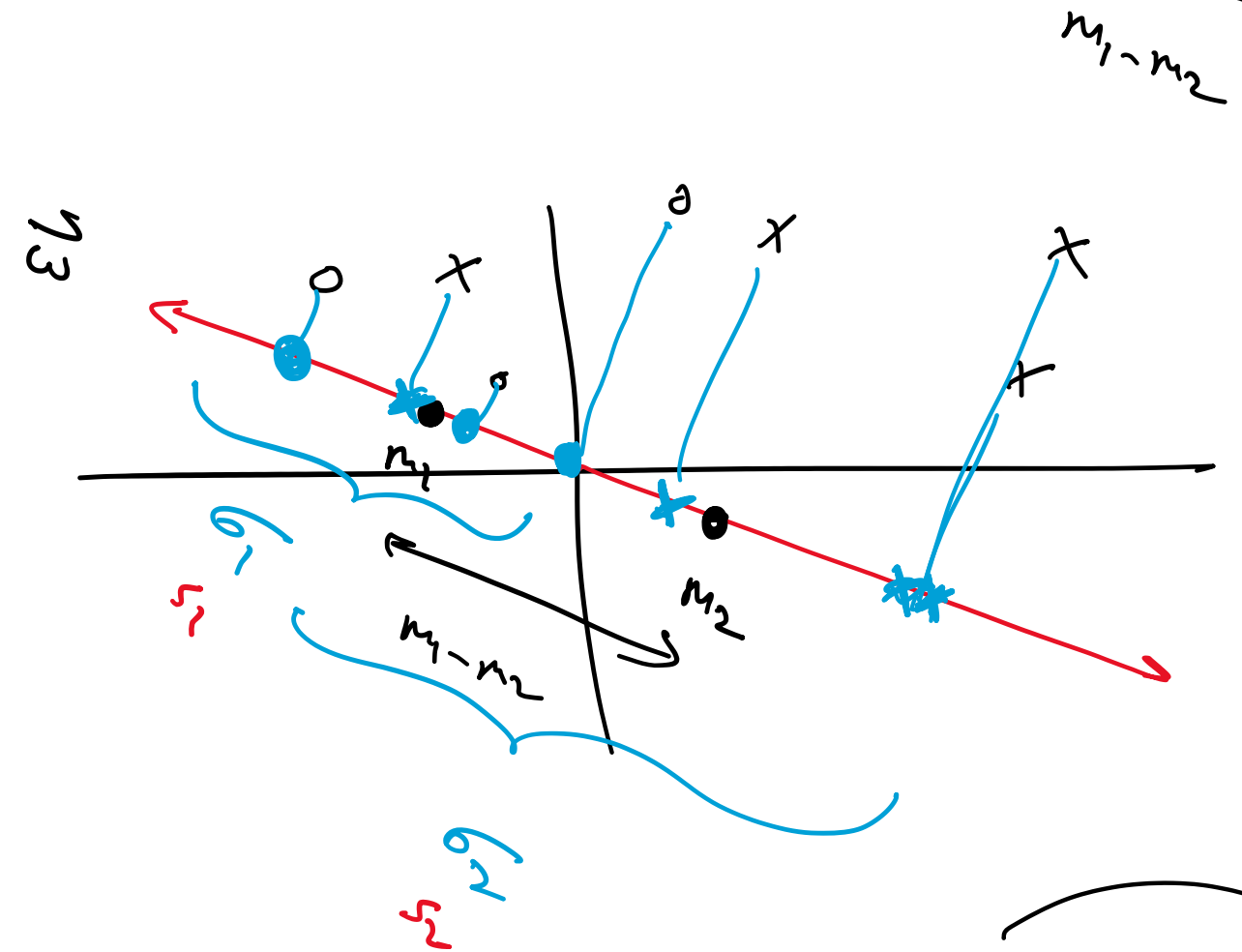
Means should be far apart $\left\{ \begin{array}{l} m_1 \text{ is mean of projected points for } c_1 \\ m_2 \quad '' \quad c_2 \end{array} \right.$

② minimize the scatter (unaveraged variance)

go for compact "groups"

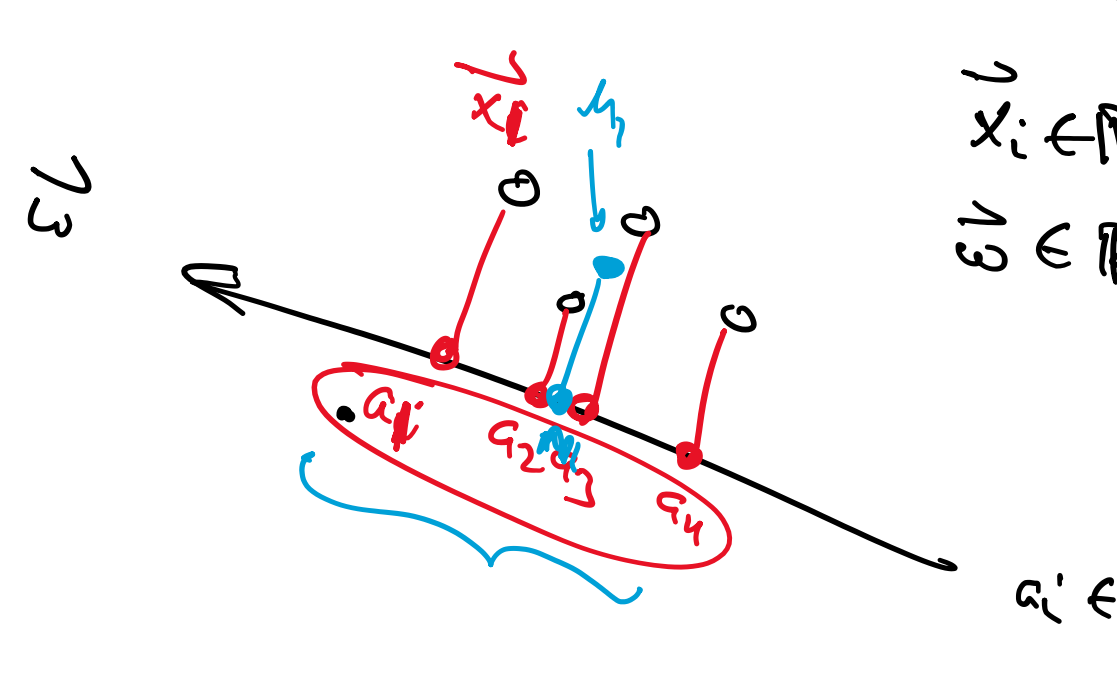
$\underbrace{s_1^2 + s_2^2}_{\text{minimize}}$





objective $\equiv \max_{\vec{\omega}} J = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \rightarrow \vec{\omega}^T B \vec{\omega}$

$\vec{\omega}$?



$\vec{x}_i \in \mathbb{R}^d$

$\vec{\omega} \in \mathbb{R}^d$

$a_i \in \mathbb{R}$

$\|\vec{\omega}\| = 1$

$\vec{\omega}$ should be a unit vector

$\vec{\omega}^T \omega = 1$

$a_i \equiv$ scalar projection

$a_i = \frac{\vec{x}_i^T \vec{\omega}}{\vec{\omega}^T \vec{\omega}} = \vec{x}_i^T \vec{\omega} = \vec{\omega}^T \vec{x}_i$

$$\sigma^2 = \frac{\sum_{a_i \in c_1} (a_i - m_1)^2}{n_1}$$

$$\boxed{s_1^2 = \sum_{a_i \in c_1} (a_i - m_1)^2}$$

class c_1

$m_1 = \frac{1}{n_1} \sum_{a_i \in c_1} a_i$

$= \frac{1}{n_1} \sum \vec{\omega}^T \vec{x}_i$

$m_1 = \vec{\omega}^T \left(\frac{\sum_{x_i \in C_1} \vec{x}_i}{n_1} \right)$

$\boxed{m_1 = \vec{\omega}^T \vec{\mu}_1}$

Projected mean   Projection of the d-dim mean

$m_2 = \vec{\omega}^T \vec{\mu}_2$

numerator

$(m_1 - m_2)^2$

$(\vec{\omega}^T \vec{\mu}_1 - \vec{\omega}^T \vec{\mu}_2)^2$

$[\vec{\omega}^T (\vec{\mu}_1 - \vec{\mu}_2)]^2$

$[\vec{\omega}^T (\vec{\mu}_1 - \vec{\mu}_2) \times (\vec{\mu}_1 - \vec{\mu}_2)^T \vec{\omega}]$

$\vec{\omega}^T \left(\underbrace{(\vec{\mu}_1 - \vec{\mu}_2)(\vec{\mu}_1 - \vec{\mu}_2)^T}_{\substack{d \times d \text{ matrix} \\ \text{(rank one)}}} \right) \vec{\omega}$

B [between cluster matrix]

$\vec{\omega}^T B \vec{\omega}$ quadratic form

$\vec{x}^T \Sigma \vec{x}$

$\vec{x}^T \Sigma^{-1} \vec{x}$

$\vec{u}^T L \vec{u}$

$B \equiv$ outer product of the difference vector

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

$\mathbb{R}^d$

$s_1^2 = \sum_{a_i \in c_1} (a_i - m_1)^2$

scatter

$= \sum_{a_i \in c_1} (\vec{\omega}^T \vec{x}_i - \vec{\omega}^T \vec{\mu}_1)^2$

$= \vec{\omega}^T \left(\underbrace{\sum_{a_i \in C_1} (\vec{x}_i - \vec{\mu}_1)(x_i - \mu_1)^T}_{} \right) \vec{\omega}$

$S_1 \equiv$ scatter matrix

$S_1 = n_1 \Sigma_1$ ← Cov matrix for class c_1

$s_1^2 = \vec{\omega}^T S_1 \vec{\omega}$

$s_2^2 = \vec{\omega}^T S_2 \vec{\omega}$

Objective: $J = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} = \frac{\vec{\omega}^T B \vec{\omega}}{\underbrace{\vec{\omega}^T S_1 \vec{\omega}}_{s_1^2} + \underbrace{\vec{\omega}^T S_2 \vec{\omega}}_{s_2^2}}$

$$\boxed{\max_{\vec{\omega}} J = \frac{\vec{\omega}^T B \vec{\omega}}{\vec{\omega}^T S \vec{\omega}}}$$

within class

$S = S_1 + S_2$

$d \times d$ sum of scatter matrices

B and $S = S_1 + S_2$

$(\vec{\mu}_1 - \vec{\mu}_2)(\vec{\mu}_1 - \vec{\mu}_2)^T$

$\max_{\vec{\omega}} J = \frac{\vec{\omega}^T B \vec{\omega}}{\vec{\omega}^T S \vec{\omega}} = \vec{\omega}^T (S^{-1} B) \vec{\omega}$

LDA $\max_{\vec{\omega}} J = \vec{\omega}^T (S^{-1} B) \vec{\omega}$ subject to $\vec{\omega}^T \vec{\omega} = 1$

$+ \lambda(\vec{\omega}^T \vec{\omega} - 1)$

PCA $\max_{\vec{u}} J_{PCA} = \vec{u}^T \Sigma \vec{u}$ subject to $\vec{u}^T u = 1$

$\frac{\partial J}{\partial \vec{\omega}} = \vec{0}$

Solution: $\vec{\omega} \Rightarrow$ dominant eigenvector of $S^{-1} B$

Q. what if there are k classes $C_1\ C_2 \ldots C_k$

C_1 vs rest — $\vec{\omega}_1$

rest → C_2, remaining — $\vec{\omega}_2$

sequence of binary problem.

---

binary case: C_1 & C_2

Solve directly (B is rank one)

$\vec{\omega}$ is the dominant eigenvector of $S^{-1}B$

$(S^{-1}B)\vec{\omega} = \lambda \vec{\omega}$ ← Eigenvector, ← eigenvalue

$S^{-1} (\mu_1 - \mu_2) \underbrace{(\mu_1 - \mu_2)^T \vec{\omega}}_{b} = \lambda \vec{\omega}$

$d \times d$ $d \times 1$

$$\boxed{\cancel{\frac{b}{\lambda}}\ \underbrace{S^{-1} (\vec{\mu}_1 - \vec{\mu}_2)} = \vec{\omega}}$$

just normalize $\vec{\omega}$ to be unit vector

$\vec{\omega}' = \vec{\omega} / \|\vec{\omega}\|$

Linear regression

$\underbrace{\vec{x}_i \in \mathbb{R}^d}_{\text{independent}}$, $\underbrace{y_i \in \mathbb{R}}_{\text{dependent/target}}$

$\hat{y} = f(A_1, A_2, \ldots, A_d) + \varepsilon$ ← noise term (Unknown attribute)

predicted value   ← what class of model?

$$\boxed{\hat{Y} = w_0 \vec{1} + w_1 \vec{A}_1 + w_2 \vec{A}_2 + \cdots + w_d \vec{A}_d}$$ ← linear model

$$\boxed{\hat{y} = \underbrace{w_0 \cdot 1}_{\text{bias}} + \underbrace{w_1 x_1 + w_2 x_2 + \cdots + w_d x_d}_{\text{weight}}}$$

$\vec{x} \in \mathbb{R}^d$

Unknown $w_0, w_1, w_2, \ldots, w_d$

$\hat{y} = \vec{\omega}^T \vec{x} + w_0$

$\vec{\omega} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix} \in \mathbb{R}^d$

$\vec{x} \in \mathbb{R}^d = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$

$\frac{\vec{x}^T \vec{1}}{\vec{1}^T \vec{1}} \rightarrow \mu$

Scalar projection onto $\vec{1}$ gives the mean value

squared error $= (y - \hat{y})^2$

true value   Predicted value

$D \ni D_{train}$ $(\vec{x}_i, y_i) \rightarrow \boxed{f} \rightarrow \hat{y}_i$

Point   ground truth   learn f $w_0, \vec{\omega} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{pmatrix}$ predicted value

$\underbrace{\min_{w_0 w_1 \ldots w_d}} J = \mathcal{L} = \sum_{i=1}^{n} \underbrace{(y_i - \hat{y}_i)^2}_{e_i} \equiv$ Sum of squared errors (SSE)

loss function "min"

error function

SSE $\equiv \mathcal{L} = \sum_{i=1}^{n} e_i^2$

$e_i = y_i - \hat{y}_i$ ← error for $\vec{x}_i$

---

d = 1

bivariate regression

A_1	Y
x_1	y_1
x_2	y_2
$\vdots$	
x_n	y_n

$\hat{y}_i = w_0 + w_1 x_i$ ← A_1

$\mathcal{L} = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$

$$\boxed{\min_{w_0, w_1} \mathcal{L} = \sum_i (y_i - w_0 - w_1 x_i)^2}$$

$\frac{\partial \mathcal{L}}{\partial w_0} = \sum_i 2(y_i - w_0 - w_1 x_i) \cdot (-1) = 0$

$\sum y_i - \sum w_0 - \sum w_1 x_i = 0$

$n \cdot w_0 = \sum y_i - w_1 \sum x_i$

$w_0 = \left(\frac{1}{n} \sum y_i\right) - w_1 \left(\frac{\sum x_i}{n}\right)$

$$\boxed{w_0 = \mu_y - w_1 \mu_{A_1}}$$ bias

A_1	Y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

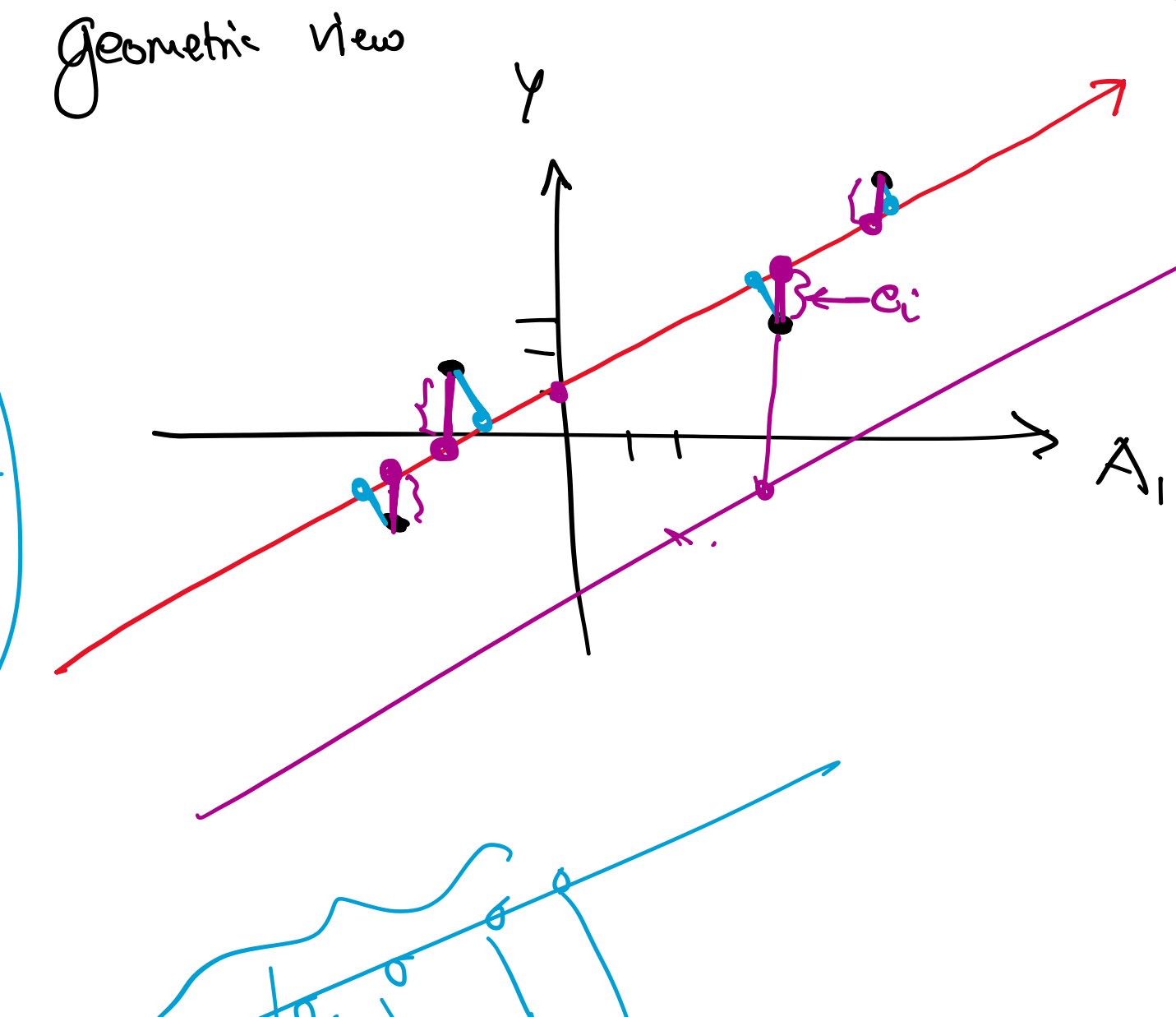
$\frac{\partial \mathcal{L}}{\partial w_1} = \sum 2(y_i - w_0 - w_1 x_i)(-x_i) = 0$

$$\boxed{w_1 = \frac{\sum x_i y_i - n \mu_x \mu_y}{\sum x_i^2 - n \cdot \mu_x^2} = \frac{Cov(X,Y)}{Var(X)} = \frac{\sigma_{XY}}{\sigma_X^2}}$$

weight

Unintuitive!

Geometric view



x_1, y_1

2, 3

PCA ≡ orthogonal projection

min SSE or MSE

only direction matters

Point view

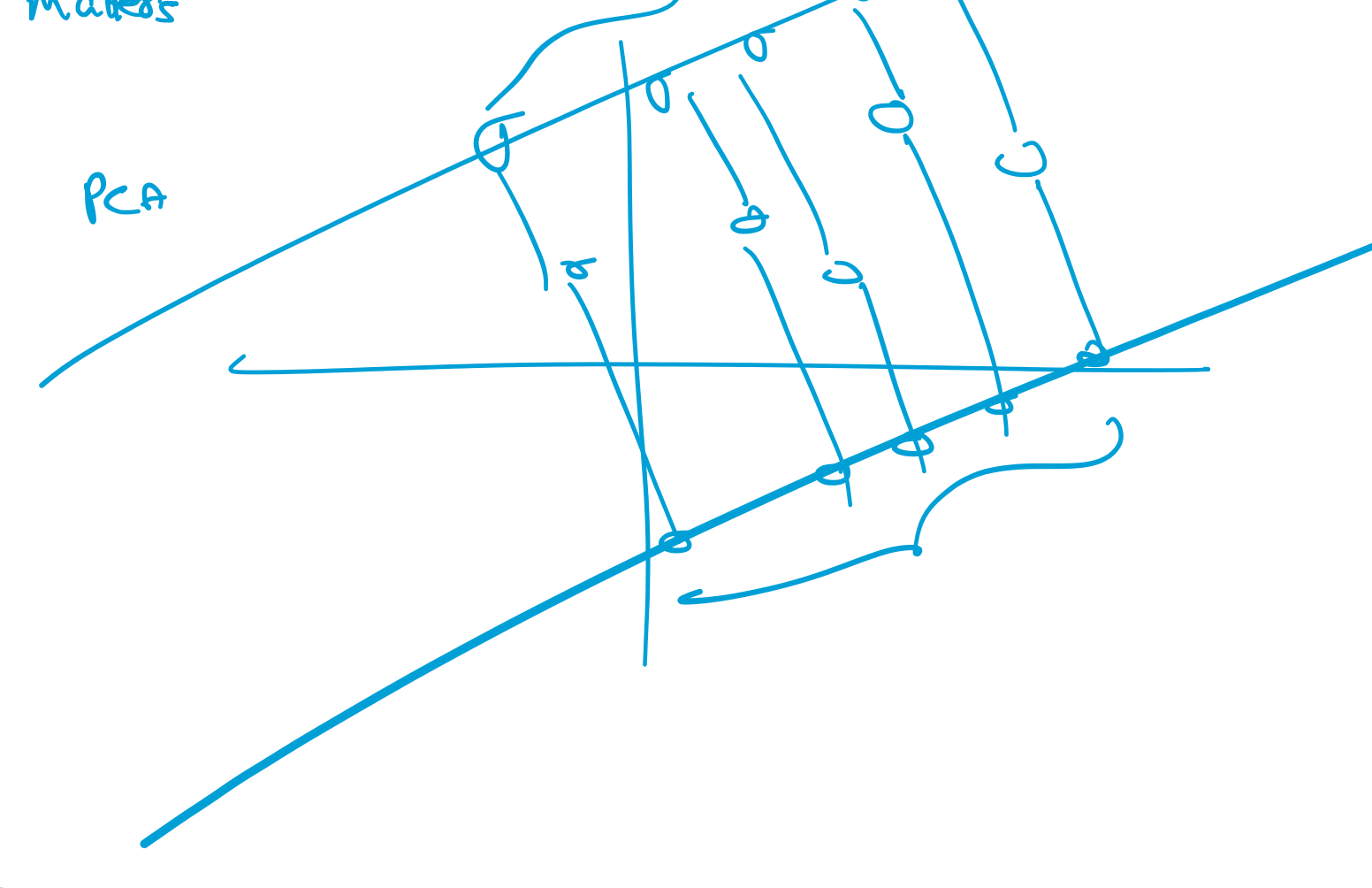
linear regression

$\sum e_i$

Vertical projection

min SSE

both the direction & the bias/offset



Column perspective

$\hat{y}_1 = w_0 \cdot 1 + w_1 x_1$

$\hat{y}_2 = w_0 \cdot 1 + w_1 x_2$

$\vdots$

$\hat{y}_n = w_0 \cdot 1 + w_1 x_n$

n equations

$\mathcal{L} = \sum (y_i - \hat{y}_i)^2$

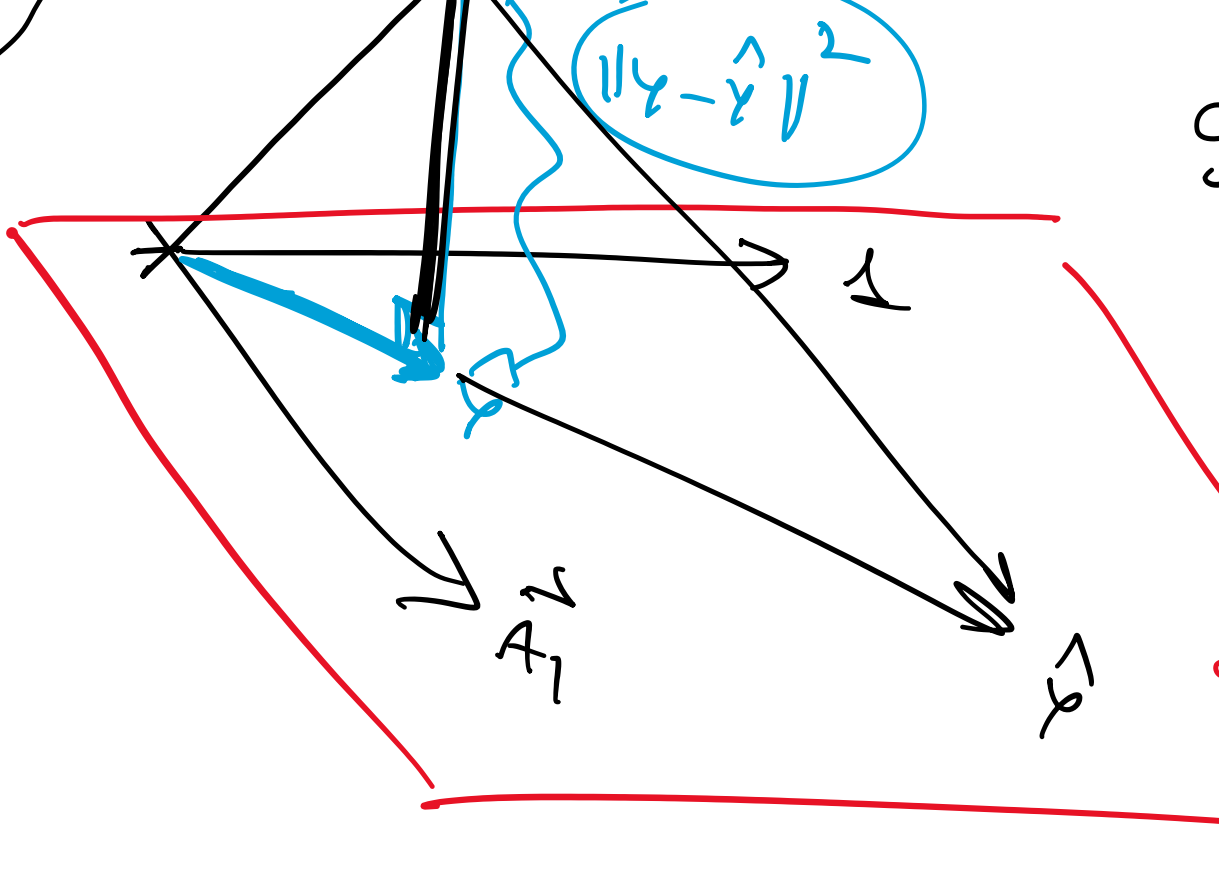
w_0, w_1

$\vec{1}$	A_1	Y	$\hat{Y}$
1	x_1	y_1	$\hat{y}_1$
1	x_2	y_2	$\hat{y}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	x_n	y_n	$\hat{y}_n$

$\underbrace{\vec{1}, \vec{A}_1}_{\text{input}}, Y, \underbrace{\hat{Y}}_{\text{output}}$ ← four n-dim vectors

$$\boxed{\hat{Y} = w_0 \cdot \vec{1} + w_1 \vec{A}_1}$$

$\hat{Y}$ should be a linear combination of $\vec{1}, \vec{A}_1$



Input comprise

$\vec{A}_1, \vec{1}, \vec{Y}$