10-4-2025

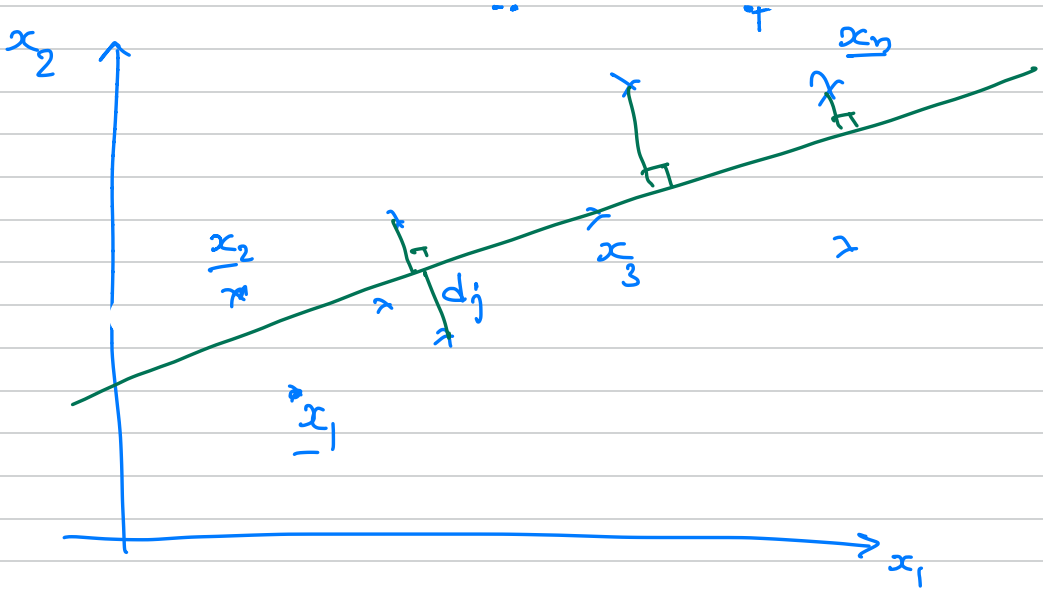
Principal Components Analysis (PCA)

$X = (X_1, \dots, X_p)$

δεδομένα

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{matrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \end{matrix}$$

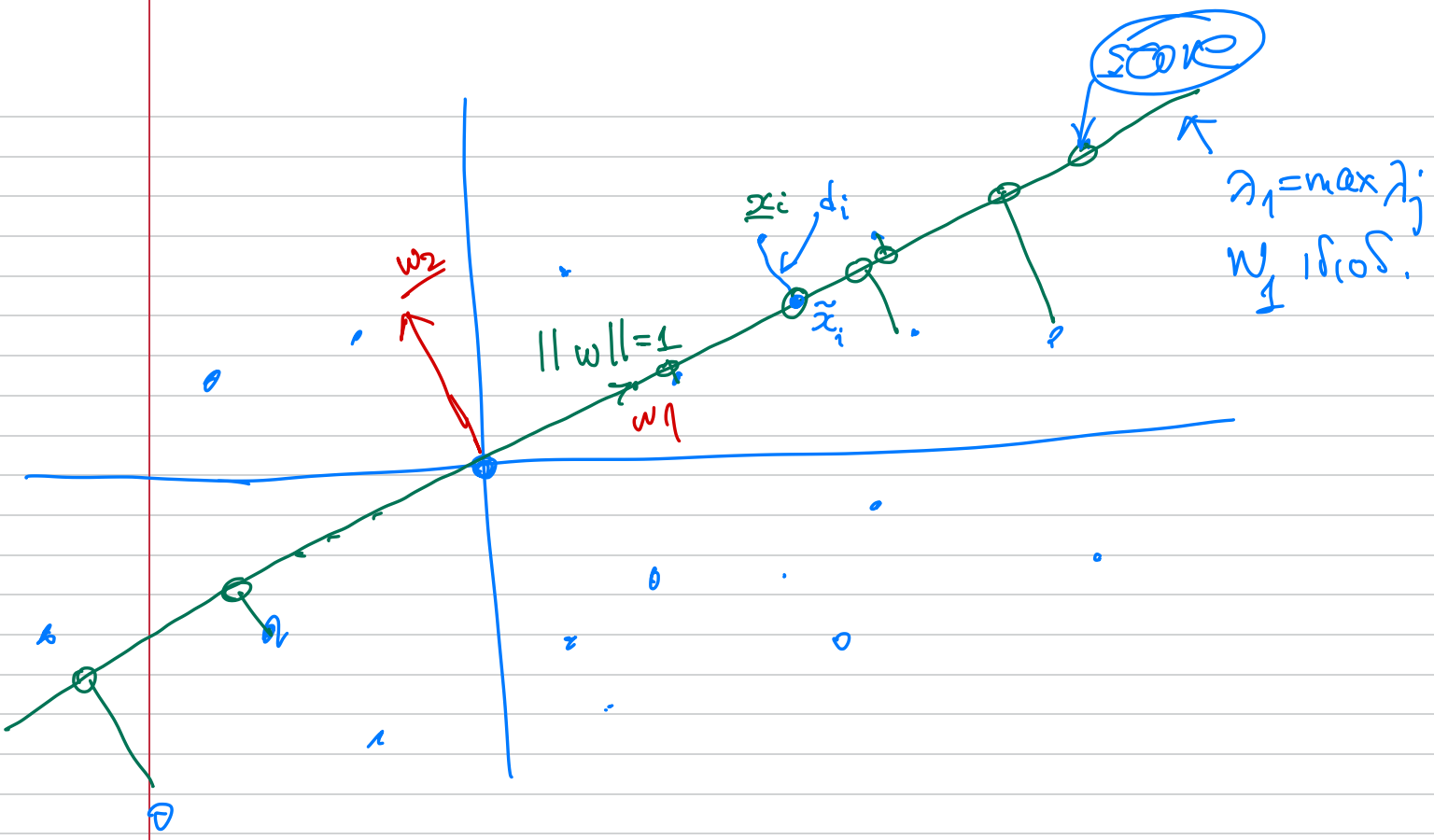
$x_{..}$



@ Κεντράρισμα : $x'_{ij} = x_{ij} - \bar{x}_{\cdot j}$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^{n} x'_{ij} = 0$$

χωρίς υποθέσεις $\frac{1}{n} \sum_{i=1}^{n} x_{ij} = 0$. $(\bar{x}_{\cdot j} = 0 \ \forall j)$



$$\tilde{x}_i = (x_i' \cdot w) \cdot w$$

$$\sum \tilde{x}_i = 0$$

$$d_i = x_i - \tilde{x}_i = x_i - (x_i^T w) \cdot w$$

$$\|d_i\|^2 = \| x_i - (x_i^T w) w \|_2^2$$

$$= (x_i - (x^T w) w)^T \cdot (x_i - (x^T w) w) =$$

$$\ldots = x_i^T \cdot x_i - (x_i^T w)^2$$

$$\min \frac{1}{\eta} \sum_{i=1}^{\eta} d_i^2 \iff \max \frac{1}{\eta} \sum_{i=1}^{\eta} (x_i^T w)^2$$

$\iff$ max $\mathrm{Var}(x^T \cdot w) \equiv \sigma_w^2$ ← διαγραπτική διασπορά της $Z = X^T \cdot W$

Επειδή υποθ. $E(X) = 0$

$X = (X_1, \ldots, X_p)$

$E(X^T \cdot w) = 0$

$\Rightarrow Var(X^T \cdot w) = E\left[(X^T \cdot w)^2\right]$

$\sigma_w^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i^T \cdot w)^2 = \frac{1}{n} (Xw)^T (Xw)$ (αφού επαληθεύτε)

$$X = \begin{pmatrix} x_{11} \cdots x_{1p} \\ \\ x_{n1} \cdots x_{np} \end{pmatrix}$$

$\sigma_w^2 = \frac{1}{n} w^T \cdot X^T X \cdot w$

Πρόβλημα $\max \; w^T V w$

υ.π. $w^T w = 1$

$\left(V = \frac{1}{n} X^T X\right)$

$$\mathcal{L} = w^T V w - \lambda (w^T w - 1)$$

$$\nabla_w \mathcal{L} = 2Vw - 2\lambda I w = 2(V - \lambda I) \cdot w$$

$$\nabla_\lambda \mathcal{L} = w^T w - 1 = 0 \Rightarrow w^T w = 1$$

$$\nabla_w \mathcal{L} = 0 \Rightarrow Vw = \lambda w \qquad V_{p \times p}$$

λ ιδιοτιμή V

w ιδιοδιάνυσμα V

$V = X^T X$ συμμετρ. θετικά ημιορ.

ιδιοτιμές $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$.

ιδιοδιανύσματα: ορθογώνια.

$$\sigma_w^2 \leftarrow \max \Rightarrow \lambda = \lambda_1 = \max \lambda_j$$

Εφαρμογή PCA στο dataset Auto.