Ex3)

linear separation with the least square principle:

$x^{(n)} \Rightarrow$ training vector , $y(n)$ class label

$\hat{y} = W^T X \Rightarrow$ linear classifer , $w_0 = 0$

* the cost / loss function

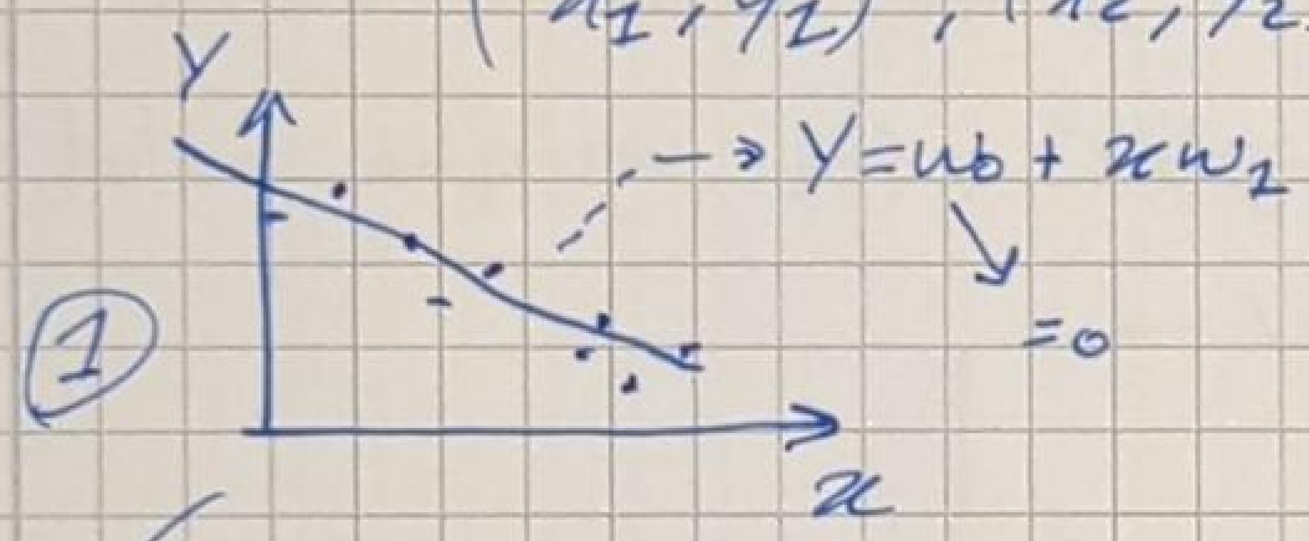
$$E(w) = \sum_{n=1}^{N} (y^{(n)} - W^T x^{(n)})^2 \quad ---- \quad ①$$

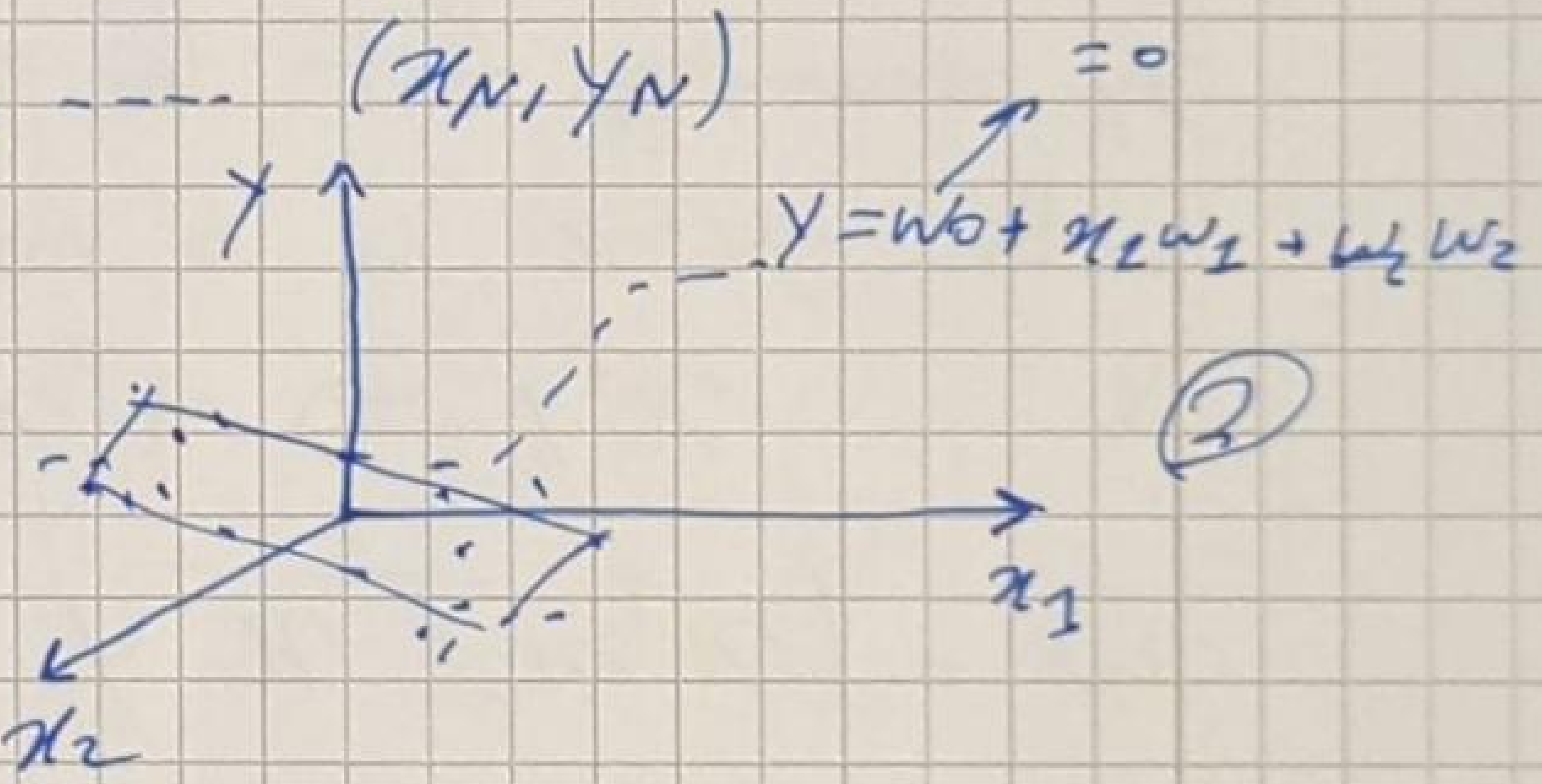
$$E(w) = (y - W^T x)(y - W^T x)^T \quad ---- \quad ②$$

ⓐ The E(w) function minimum is always exists but not unique. The training vectors represent independent random draws. Therefore, the X represents a $N \times P$ matrix with each row an input vector and y is an N-vector of the outputs in the training set.

The ~~output~~ in of observation pairs:

$$(x_1, y_1), (x_2, y_2), ---- (x_N, y_N)$$





the point in ① lies in (x_p, y_p) of the slope w_1 as in the two dimensions as $w_0 + x w_1 = y$ is close to $w_0 + x_p w_1 = y$

the point in ③ a data of three dimensions is defind as $w_0 + x_1 w_1 + x_2 w_2 + \dots + x_N w_N = y$ ~~the~~ which holds to a hyperplane coordinate slop of $N=2$

By presenting the linear models and least squares and giving a X^T as an input vector $X^T = (X_1, X_2, \dots X_p)$ the summation would be used as:

$$\hat{Y} = \hat{w}_0 + \sum_{n=1}^{N} X^{(n)} \hat{w}$$